

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 14:02:19 ; Search time 3422 seconds
(without alignments)
18755.907 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 3963
Sequence: 1 atcgacgacagcataga.....tggagagtgatgaactttaa 3963

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estcom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	478	12.1	553	12	BF126187
2	291	7.3	931	14	BQ218668
3	270	6.8	503	17	AQ621045
4	212	5.3	481	9	AA179599
5	165	4.2	331	17	AQ167332
6	159	4.0	439	17	AQ768080

c 80	20	0.5	328	9	AA305887	AA305887 EST176886	c 153	20	0.5	432	14	N37003	N37003 yy40a03.s1
c 81	20	0.5	330	9	AA143537	AA143537 zo32b10.s	c 154	20	0.5	434	9	AA171726	AA171726 zo95e01.s
c 82	20	0.5	330	10	AW472766	AW472766 xq19b11.x	c 155	20	0.5	434	10	AW591744	AW591744 xc87g10.x
c 83	20	0.5	331	14	T05375	T05375 EST03264 Fe	c 156	20	0.5	435	14	BM825747	BM825747 K-EST0097
c 84	20	0.5	331	9	AA758826	AA758826 ah76d05.s	c 157	20	0.5	435	9	AA113239	AA113239 zn66d05.s
c 85	20	0.5	331	10	BB550055	BB550055 BB550055	c 158	20	0.5	437	12	BG225506	BG225506 kp68f10.y
c 86	20	0.5	333	9	AA070066	AA070066 zm60d04.r	c 159	20	0.5	437	14	W43020	W43020 zc84e11.s1
c 87	20	0.5	333	9	AA634124	AA634124 ac34e03.s	c 160	20	0.5	437	14	W43020	W43020 zc84f11.s1
c 88	20	0.5	334	10	BB519999	BB519999 BB519999	c 161	20	0.5	438	9	AA399590	AA399590 zt93d04.s
c 89	20	0.5	334	14	W80897	W80897 z88e06.r1	c 162	20	0.5	439	10	AW275274	AW275274 xv79a12.x
c 90	20	0.5	336	9	AA070835	AA070835 zm60d04.s	c 163	20	0.5	441	14	N20632	N20632 yx40a05.s1
c 91	20	0.5	338	9	AI687776	AI687776 tp03c12.x	c 164	20	0.5	442	9	AA457447	AA457447 aa86d09.r
c 92	20	0.5	341	9	AI581154	AI581154 t194b03.x	c 165	20	0.5	442	9	AA578597	AA578597 nk92h06.s
c 93	20	0.5	342	9	AI268504	AI268504 qo37c06.x	c 166	20	0.5	442	14	W60512	W60512 zc99e01.s1
c 94	20	0.5	342	17	AZ949697	AZ949697 2M0213N03	c 167	20	0.5	443	14	H80208	H80208 yu56h07.r1
c 95	20	0.5	346	14	F10573	F10573 HSC3HA112 n	c 168	20	0.5	443	14	W42769	W42769 zc84g02.s1
c 96	20	0.5	349	14	G68587	G68587 yH99f04.s1	c 169	20	0.5	444	9	AI199880	AI199880 qf84g08.x
c 97	20	0.5	353	14	BM999157	BM999157 UI-H-DI0-	c 170	20	0.5	444	13	BG926835	BG926835 HNC24-1-H
c 98	20	0.5	354	14	N26423	N26423 yx23a06.r1	c 171	20	0.5	444	14	H11623	H11623 ym17a02.s1
c 99	20	0.5	357	9	AI864587	AI864587 wl54f09.x	c 172	20	0.5	445	9	AI423339	AI423339 tf36c06.x
c 100	20	0.5	360	9	AI365397	AI365397 qz08d01.x	c 173	20	0.5	445	14	BM758152	BM758152 K-EST0037
c 101	20	0.5	361	9	AI721077	AI721077 as69g07.x	c 174	20	0.5	447	9	AI214594	AI214594 qm28h03.x
c 102	20	0.5	362	14	N67089	N67089 yz45f04.s1	c 175	20	0.5	447	14	R61550	R61550 ym16g04.s1
c 103	20	0.5	364	12	BG475893	BG475893 mac03f09.	c 176	20	0.5	447	14	W42768	W42768 zc84g01.s1
c 104	20	0.5	366	12	BG054566	BG054566 7f18g04.x	c 177	20	0.5	448	9	AI632835	AI632835 tz32g10.x
c 105	20	0.5	367	14	H13759	H13759 yf09g01.r1	c 178	20	0.5	449	9	AA412702	AA412702 zu12g07.s
c 106	20	0.5	369	9	AI700275	AI700275 wd05g01.x	c 179	20	0.5	449	9	AA412703	AA412703 zu12g08.s
c 107	20	0.5	370	12	BF896448	BF896448 RC2-WT014	c 180	20	0.5	449	14	N23217	N23217 yx70d12.s1
c 108	20	0.5	370	12	BF896482	BF896482 RC2-WT014	c 181	20	0.5	451	10	AW265015	AW265015 xq59h02.x
c 109	20	0.5	372	9	AI020253	AI020253 ow72d07.x	c 182	20	0.5	454	14	BM761438	BM761438 K-EST0042
c 110	20	0.5	372	10	AW296975	AW296975 UI-H-BW0-	c 183	20	0.5	455	12	BG179037	BG179037 602330745
c 111	20	0.5	375	12	BF843256	BF843256 PM0-HT107	c 184	20	0.5	456	9	AI129577	AI129577 qc40c08.x
c 112	20	0.5	376	14	W05302	W05302 za83h01.r1	c 185	20	0.5	458	10	AV713167	AV713167 yu75b07.r1
c 113	20	0.5	377	9	AA857896	AA857896 oe85d02.s	c 186	20	0.5	459	14	H05956	H05956 yl76d02.s1
c 114	20	0.5	378	9	AI152320	AI152320 ud87h11.r	c 187	20	0.5	460	9	AI245353	AI245353 qx28g12.x
c 115	20	0.5	381	10	AW244099	AW244099 xs50h10.x	c 188	20	0.5	461	9	AA725804	AA725804 ai23d08.s
c 116	20	0.5	381	10	AA000249	AA000249 zu62e09.s	c 189	20	0.5	462	10	AW102901	AW102901 xd40a09.x
c 117	20	0.5	381	10	AAW64743	AAW64743 he19a05.x	c 190	20	0.5	464	9	AI815029	AI815029 wk70b09.x
c 118	20	0.5	383	9	AA770006	AA770006 ah81g06.s	c 191	20	0.5	464	14	H94909	H94909 yu57b07.r1
c 119	20	0.5	383	9	AI025877	AI025877 ow12e10.s	c 192	20	0.5	465	9	AI539659	AI539659 tp60h10.x
c 120	20	0.5	383	14	BM761821	BM761821 K-EST0042	c 193	20	0.5	465	13	BM664112	BM664112 UI-E-CK1-
c 121	20	0.5	384	14	C74987	C74987 C74987 Huma	c 194	20	0.5	465	14	H63616	H63616 yr26g02.r1
c 122	20	0.5	384	14	W02651	W02651 zc64h10.s1	c 195	20	0.5	466	9	AA503401	AA503401 ne45h11.s
c 123	20	0.5	395	9	AA159427	AA159427 zo61g08.r	c 196	20	0.5	467	9	AA405085	AA405085 zu12g07.r
c 124	20	0.5	396	9	AI559943	AI559943 tq77a09.x	c 197	20	0.5	467	14	N28783	N28783 yx69h12.r1
c 125	20	0.5	398	10	AW136170	AW136170 UI-H-B11-	c 198	20	0.5	467	17	AQ126297	AQ126297 HS.3036.B
c 126	20	0.5	400	9	AI843050	AI843050 UI-M-AK1-	c 199	20	0.5	468	14	H03624	H03624 yj37b09.r1
c 127	20	0.5	403	14	BQ219817	BQ219817 AGENCOURT	c 200	20	0.5	469	14	R44291	R44291 y934b12.s1
c 128	20	0.5	404	14	R39327	R39327 yd01f07.s1	c 201	20	0.5	470	9	AA644539	AA644539 abe2g08.s
c 129	20	0.5	406	10	AW572448	AW572448 xq18b11.x	c 202	20	0.5	470	9	AI742544	AI742544 wb55c01.x
c 130	20	0.5	406	12	BF382888	BF382888 601816711	c 203	20	0.5	471	9	AA398996	AA398996 zt93d04.r
c 131	20	0.5	407	12	BF988650	BF988650 QV1-WT022	c 204	20	0.5	471	9	AA558422	AA558422 nk38c01.s
c 132	20	0.5	408	9	AA143786	AA143786 zo31c02.s	c 205	20	0.5	473	9	AA948026	AA948026 oq58h12.s
c 133	20	0.5	408	10	AW238387	AW238387 xp24f03.x	c 206	20	0.5	475	9	AA843760	AA843760 aj18b12.s
c 134	20	0.5	409	9	AA903459	AA903459 ok58g01.s	c 207	20	0.5	475	14	C06009	C06009 C06009 Huma
c 135	20	0.5	411	9	AI267718	AI267718 ag93h06.x	c 208	20	0.5	476	9	AI829320	AI829320 wk58g07.x
c 136	20	0.5	412	9	AA854663	AA854663 aj63g05.s	c 209	20	0.5	479	9	AI820088	AI820088 wq56b09.x
c 137	20	0.5	414	14	H13707	H13707 yj09g01.s1	c 210	20	0.5	479	9	AA526754	AA526754 nl56a12.s
c 138	20	0.5	414	12	BE762819	BE762819 QV3-WT002	c 211	20	0.5	480	10	BE393233	BE393233 601306438
c 139	20	0.5	414	14	T96399	T96399 ye34c05.s1	c 212	20	0.5	482	10	BE200481	BE200481 ug63e06.x
c 140	20	0.5	417	9	AA732754	AA732754 ai28c10.s	c 213	20	0.5	482	14	C06159	C06159 C06159 Huma
c 141	20	0.5	419	14	H94852	H94852 yu57b07.s1	c 214	20	0.5	483	14	N30039	N30039 yw80c11.s1
c 142	20	0.5	421	9	AA255029	AA255029 mz85b12.r	c 215	20	0.5	484	10	AW009816	AW009816 ws87h03.x
c 143	20	0.5	423	10	AW377536	AW377536 QV0-CT022	c 216	20	0.5	485	9	AI775949	AI775949 EST257049
c 144	20	0.5	424	9	AI335353	AI335353 qo88d06.x	c 217	20	0.5	489	14	BM840929	BM840929 K-EST0118
c 145	20	0.5	424	14	T52636	T52636 ya72a08.s1	c 218	20	0.5	491	9	AA151703	AA151703 z139f04.s
c 146	20	0.5	426	9	AI826788	AI826788 wk56c06.x	c 219	20	0.5	491	14	BM793241	BM793241 K-EST0073
c 147	20	0.5	427	10	AW236394	AW236394 xol5e11.x	c 220	20	0.5	491	14	T32456	T32456 EST48711 Hu
c 148	20	0.5	429	17	BM115735	BM115735 RPCI-24-2	c 221	20	0.5	493	14	BM968955	BM968955 UI-CF-DU1
c 149	20	0.5	431	10	AW470683	AW470683 ha31c11.x	c 222	20	0.5	494	14	AA166828	AA166828 zq39f01.s
c 150	20	0.5	431	9	AA311653	AA311653 EST182513	c 223	20	0.5	494	14	BM841252	BM841252 K-EST0118
c 151	20	0.5	431	10	BE246590	BE246590 TCBAP1E49	c 224	20	0.5	495	9	AA126150	AA126150 z186e05.s
c 152	20	0.5	432	12	BG089773	BG089773 mab80h03..	c 225	20	0.5	499	10	AW513554	AW513554 xo86c11.x

372	19	0.5	187	14	N88589	N88589_K3927F_Huma	445	19	0.5	420	9	AJ476027	AJ476027
c 373	19	0.5	189	12	BG005404	PM1-GN018	446	19	0.5	420	10	AW049514	AW049514
c 374	19	0.5	200	14	F13731	ATTS4726_Gi	c 447	19	0.5	421	14	H65241	H65241
c 375	19	0.5	201	17	AZ285588	RPCI-23-1	c 448	19	0.5	422	9	AA292934	AA292934
c 376	19	0.5	202	10	AW391879	QV4-ST023	c 449	19	0.5	425	13	BM276362	BM276362
c 377	19	0.5	212	13	BI335744	16d182_Hu	450	19	0.5	427	12	BG010693	BG010693
c 378	19	0.5	218	10	B8217941	B8217941	451	19	0.5	428	14	BM839592	BM839592
c 379	19	0.5	215	17	AZ929941	AZ929941_479_G1156	c 452	19	0.5	429	14	BQ335411	BQ335411
c 380	19	0.5	227	10	AW176118	CM0-BT015	453	19	0.5	435	13	BI085464	BI085464
c 381	19	0.5	239	10	BB147979	BB147979	454	19	0.5	438	9	AA858955	AA858955
c 382	19	0.5	243	10	BE525580	M62A19STM	c 455	19	0.5	438	12	BE819588	BE819588
c 383	19	0.5	246	12	BF330449	BF330449	c 456	19	0.5	438	14	W43025	W43025
c 384	19	0.5	272	10	AV373415	AV373415	c 457	19	0.5	441	9	AI124100	AI124100
c 385	19	0.5	282	12	BF057414	7k43612_x	458	19	0.5	442	10	BE517044	BE517044
c 386	19	0.5	294	12	BF910144	CM2-UT008	459	19	0.5	442	14	R72980	R72980
c 387	19	0.5	302	9	AV161533	AV161533	460	19	0.5	443	10	BE544732	BE544732
c 388	19	0.5	309	12	BE839410	BE839410_RC3-FN014	c 461	19	0.5	444	9	AI559477	AI559477
c 389	19	0.5	312	14	BQ802199	WHE2823_B	c 462	19	0.5	444	17	AQ613337	AQ613337
c 390	19	0.5	313	10	BE416182	MUG005_E0	c 463	19	0.5	446	9	AI967571	AI967571
c 391	19	0.5	313	10	BE417526	BE417526_MUG021_H0	c 464	19	0.5	447	9	AA132343	AA132343
c 392	19	0.5	317	14	BQ107249	NXLV090_A	c 465	19	0.5	450	14	BQ010859	BQ010859
c 393	19	0.5	318	12	BF760287	PM4-CT064	c 466	19	0.5	450	14	BQ013790	BQ013790
c 394	19	0.5	325	12	BF371522	RC3-FN014	c 467	19	0.5	451	9	AI810140	AI810140
c 395	19	0.5	326	14	H93947	H93947_YV13B09.r1	c 468	19	0.5	452	9	AA291984	AA291984
c 396	19	0.5	329	17	AQ776523	HS-5572_A	c 469	19	0.5	455	9	AI509009	AI509009
c 397	19	0.5	330	14	BQ355591	MR4-HT105	c 470	19	0.5	456	17	AZ742633	AZ742633
c 398	19	0.5	331	14	C97031	C97031_C97031_R1ce	471	19	0.5	459	13	BM018162	BM018162
c 399	19	0.5	333	10	AW280244	fj37a10_Y	472	19	0.5	460	14	R64455	R64455
c 400	19	0.5	338	17	BH628698	BH628698_1007084E0	473	19	0.5	466	14	R64455	R64455
c 401	19	0.5	349	9	A1271741	q177a12_x	c 474	19	0.5	466	12	BF642817	BF642817
c 402	19	0.5	349	17	AQ166227	HS-3076_B	475	19	0.5	466	13	BI981527	BI981527
c 403	19	0.5	350	12	BF365700	QV2-NT004</							

518	19	0.5	518	13	BI085469	BI085469 602869961	c 591	19	0.5	618	9	AA410991	AA410991 zv03a10.s
519	19	0.5	519	14	HI8557	HI8557 ym43d05.r1	c 592	19	0.5	618	17	AZ464702	AZ464702 IM0274008
520	19	0.5	520	17	AQ188298	AQ188298 HS-3237.A	c 593	19	0.5	619	17	AG131039	AG131039 Pan trogl
521	19	0.5	521	17	AQ625532	AQ625532 CITBI-EI-	c 594	19	0.5	622	17	AZ633587	AZ633587 IM0488J24
522	19	0.5	522	9	AI651864	AI651864 wB50906.x	c 595	19	0.5	624	12	BG676137	BG676137 602622534
523	19	0.5	523	13	BI816996	BI816996 UMN07B01	c 596	19	0.5	630	13	BM290756	BM290756 EST577290
524	19	0.5	524	10	AV566804	AV566804 AV566804	c 597	19	0.5	631	17	AZ461188	AZ461188 IM0266N22
525	19	0.5	525	12	BG666831	BG666831 DRACKEO7	c 598	19	0.5	631	17	BH210310	BH210310 Sml-58L10
526	19	0.5	526	17	AQ611342	AQ611342 HS-5088.B	c 599	19	0.5	633	12	BG675764	BG675764 602621991
527	19	0.5	527	17	AZ827736	AZ827736 2M0104P18	c 600	19	0.5	634	9	AU133678	AU133678 AU133678
528	19	0.5	528	10	AW403729	AW403729 UI-HF-BK0	c 601	19	0.5	634	13	BM018305	BM018305 603645961
529	19	0.5	529	13	BM017565	BM017565 603644741	c 602	19	0.5	634	17	AZ447165	AZ447165 IM0243119
530	19	0.5	530	9	AI249796	AI249796 qX51407.x	c 603	19	0.5	635	17	AZ452262	AZ452262 IM0252H02
531	19	0.5	531	9	AA3398950	AA3398950 zt49408.r	c 604	19	0.5	637	9	AA573374	AA573374 nk98h06.s
532	19	0.5	532	12	BF059606	BF059606 7k63d07.x	c 605	19	0.5	637	10	AV349355	AV349355 AV349355
533	19	0.5	533	12	BG672378	BG672378 DRNGCD09	c 606	19	0.5	637	10	AV832799	AV832799 AV832799
534	19	0.5	534	9	AA429894	AA429894 zw66B04.s	c 607	19	0.5	637	13	BJ483997	BJ483997 BU483997
535	19	0.5	535	10	AW501570	AW501570 UI-HF-BP0	c 608	19	0.5	642	14	BQ007582	BQ007582 UI-H-ED1-
536	19	0.5	536	17	BH281312	BH281312 CH230-38N	c 609	19	0.5	643	17	BH210319	BH210319 Sml-58L10
537	19	0.5	537	13	BM017483	BM017483 603644604	c 610	19	0.5	648	9	AI911264	AI911264 wq33e10.x
538	19	0.5	538	12	BF257366	BF257366 HVSMEF001	c 611	19	0.5	649	9	AJ396327	AJ396327 AJ396327
539	19	0.5	539	9	AU023911	AU023911 AU023911	c 612	19	0.5	649	13	BM018503	BM018503 603646331
540	19	0.5	539	12	BF434653	BF434653 7p02e10.x	c 613	19	0.5	650	12	BG709312	BG709312 602673326
541	19	0.5	540	13	BG977838	BG977838 MR2-C1012	c 614	19	0.5	650	13	BM049541	BM049541 603623482
542	19	0.5	544	13	BI868014	BI868014 ft66g09.y	c 615	19	0.5	650	17	BH285147	BH285147 CH230-39J
543	19	0.5	545	17	AQ721499	AQ721499 HS-5552.B	c 616	19	0.5	651	13	BM017771	BM017771 603645194
544	19	0.5	549	13	BM016752	BM016752 603643224	c 617	19	0.5	652	13	BM020804	BM020804 603644454
545	19	0.5	552	14	BQ336080	BQ336080 PM0-MT043	c 618	19	0.5	653	13	BI085269	BI085269 602870308
546	19	0.5	552	17	AQ669972	AQ669972 HS-5388.A	c 619	19	0.5	653	13	BM016580	BM016580 603642909
547	19	0.5	555	10	AW873503	AW873503 hc60h02.x	c 620	19	0.5	653	13	BM020807	BM020807 603644457
548	19	0.5	555	13	BI085186	BI085186 602870575	c 621	19	0.5	655	9	AL682177	AL682177 AL682177
549	19	0.5	555	17	AZ404731	AZ404731 1M0173L07	c 622	19	0.5	655	13	BI969792	BI969792 GM830009A
550	19	0.5	559	12	BF224139	BF224139 7q84a10.x	c 623	19	0.5	656	17	AZ897857	AZ897857 RPCI-24-2
551	19	0.5	561	14	BQ791252	BQ791252 E4127.Chi	c 624	19	0.5	658	12	BG167189	BG167189 602345324
552	19	0.5	561	17	AZ622118	AZ622118 1M0455L10	c 625	19	0.5	660	13	BM018021	BM018021 603645563
553	19	0.5	562	12	BG696554	BG696554 602658889	c 626	19	0.5	662	17	AZ574057	AZ574057 324PvD05
554	19	0.5	565	17	AZ767436	AZ767436 1M0566B12	c 627	19	0.5	663	12	BG676859	BG676859 602623461
555	19	0.5	566	13	BM020532	BM020532 603649262	c 628	19	0.5	663	13	BM020528	BM020528 603649257
556	19	0.5	569	10	AV947010	AV947010 AV947010	c 629	19	0.5	664	10	BB624745	BB624745 BB624745
557	19	0.5	569	13	BM019275	BM019275 603647376	c 630	19	0.5	667	12	BG677813	BG677813 602625492
558	19	0.5	569	13	BM554137	BM554137 AGENCOURT	c 631	19	0.5	668	13	BM016807	BM016807 603643286
559	19	0.5	569	17	AQ511922	AQ511922 HS-5127.A	c 632	19	0.5	668	13	BM020740	BM020740 603644261
560	19	0.5	570	10	AW852506	AW852506 PM3-CT024	c 633	19	0.5	668	13	BM049506	BM049506 603623444
561	19	0.5	571	17	AZ744481	AZ744481 RPCI-24-7	c 634	19	0.5	668	17	BH094186	BH094186 RPCI-24-2
562	19	0.5	572	14	W19673	W19673 zb36a06.r1	c 635	19	0.5	669	9	AU137100	AU137100 AU137100
563	19	0.5	573	12	BG108092	BG108092 602280145	c 636	19	0.5	669	12	BG675822	BG675822 602621861
564	19	0.5	575	9	AI595950	AI595950 uk20f11.x	c 637	19	0.5	669	14	BU036945	BU036945 946131A03
565	19	0.5	579	13	BM343810	BM343810 rr43c03.y	c 638	19	0.5	670	12	BG681956	BG681956 602629902
566	19	0.5	580	12	BF259025	BF259025 HVSMEF001	c 639	19	0.5	672	9	AL633398	AL633398 AL633398
567	19	0.5	580	13	BM016642	BM016642 603642988	c 640	19	0.5	673	10	BB197831	BB197831 BB197831
568	19	0.5	580	13	BM020892	BM020892 603648311	c 641	19	0.5	674	13	BM049538	BM049538 603623479
569	19	0.5	582	13	BM016608	BM016608 603642944	c 642	19	0.5	675	12	BG678568	BG678568 602624248
570	19	0.5	583	13	BI085523	BI085523 602870231	c 643	19	0.5	675	13	BM017453	BM017453 603644558
571	19	0.5	583	13	BM045438	BM045438 603623554	c 644	19	0.5	676	13	BM636714	BM636714 170006875
572	19	0.5	583	17	BH5611225	BH5611225 BOGSL13Tr	c 645	19	0.5	678	10	AV724825	AV724825 AV724825
573	19	0.5	587	12	BG799692	BG799692 fp12e12.y	c 646	19	0.5	679	10	AV686595	AV686595 AV686595
574	19	0.5	591	13	BM020223	BM020223 603648813	c 647	19	0.5	679	17	BH886338	BH886338 LB00726a.
575	19	0.5	592	12	BG681621	BG681621 602627920	c 648	19	0.5	680	10	BB239528	BB239528 BB239528
576	19	0.5	593	13	BI241312	BI241312 REB346.5	c 649	19	0.5	680	17	AG062467	AG062467 Pan trogl
577	19	0.5	596	13	BM017848	BM017848 603645321	c 650	19	0.5	682	12	BG436071	BG436071 602508858
578	19	0.5	597	10	AW180851	AW180851 Mga1019f	c 651	19	0.5	682	17	BH362134	BH362134 CH230-203
579	19	0.5	598	17	AZ022407	AZ022407 RPCI-23-3	c 652	19	0.5	684	10	BB629362	BB629362 BB629362
580	19	0.5	600	13	BM020723	BM020723 603644230	c 653	19	0.5	685	10	AW969104	AW969104 EST381181
581	19	0.5	602	9	AL042341	AL042341 DKFZp434P	c 654	19	0.5	685	10	BE650294	BE650294 UI-M-BH2.
582	19	0.5	603	13	BI085313	BI085313 602870371	c 655	19	0.5	685	13	BI085083	BI085083 602870619
583	19	0.5	606	9	AI400828	AI400828 th10h01.x	c 656	19	0.5	686	17	BH274863	BH274863 CH230-701
584	19	0.5	606	13	AW058699	AW058699 Fe50d01.y	c 657	19	0.5	687	12	BG676482	BG676482 602622941
585	19	0.5	606	13	BM017698	BM017698 603645050	c 658	19	0.5	688	12	BG676142	BG676142 602623539
586	19	0.5	608	13	BM045141	BM045141 603623037	c 659	19	0.5	688	12	BG677209	BG677209 602623903
587	19	0.5	609	13	BI085119	BI085119 602870665	c 660	19	0.5	688	12	BG743467	BG743467 602634986
588	19	0.5	610	13	BM020625	BM020625 603643008	c 661	19	0.5	689	10	BE293048	BE293048 601106037
589	19	0.5	612	14	BM795050	BM795050 K-EST0076	c 662	19	0.5	689	14	BQ199777	BQ199777 UI-R-DQ1-
590	19	0.5	615	13	BI866952	BI866952 ft66g09.x	c 663	19	0.5	691	12	BG678505	BG678505 602626176

c 810	19	0.5	859	12	BE737777	BE737777	601572393	883	19	0.5	1031	13	BM563005	BM563005	AGENCOURT
811	19	0.5	859	13	BI085339	BI085339	602870005	884	19	0.5	1032	13	BM017344	BM017344	603644020
812	19	0.5	862	13	BM019581	BM019581	603647812	885	19	0.5	1037	13	BM563124	BM563124	AGENCOURT
813	19	0.5	863	13	BI085402	BI085402	602870075	886	19	0.5	1037	14	BM811579	BM811579	AGENCOURT
c 814	19	0.5	867	12	BG248597	BG248597	602400767	887	19	0.5	1037	14	BM910516	BM910516	AGENCOURT
815	19	0.5	870	13	BI155748	BI155748	602904305	888	19	0.5	1038	17	CNS0423C	AL313761	Tetraodon
816	19	0.5	873	13	BM019073	BM019073	603647084	889	19	0.5	1039	13	BM552499	BM552499	AGENCOURT
817	19	0.5	876	13	BM017033	BM017033	603643582	890	19	0.5	1039	13	BM552651	BM552651	AGENCOURT
c 818	19	0.5	878	12	BE962028	BE962028	601655273	891	19	0.5	1039	14	BM908459	BM908459	AGENCOURT
819	19	0.5	881	12	BG619719	BG619719	602619258	892	19	0.5	1039	14	BM909979	BM909979	AGENCOURT
820	19	0.5	883	13	BM044961	BM044961	603622782	893	19	0.5	1043	14	BM811503	BM811503	AGENCOURT
c 821	19	0.5	884	12	BG676551	BG676551	602623022	894	19	0.5	1045	14	BM909663	BM909663	AGENCOURT
822	19	0.5	884	12	BG682700	BG682700	602629408	895	19	0.5	1046	13	BM549827	BM549827	AGENCOURT
c 823	19	0.5	886	12	BF239213	BF239213	601905422	896	19	0.5	1046	14	BM910445	BM910445	AGENCOURT
824	19	0.5	888	13	BM020008	BM020008	603648536	897	19	0.5	1048	14	BM910428	BM910428	AGENCOURT
c 825	19	0.5	888	14	BQ438967	BQ438967	AGENCOURT	898	19	0.5	1052	13	BM563350	BM563350	AGENCOURT
826	19	0.5	892	14	BQ213497	BQ213497	AGENCOURT	899	19	0.5	1053	13	BM562987	BM562987	AGENCOURT
827	19	0.5	895	13	BM019790	BM019790	603648120	900	19	0.5	1054	13	BM563407	BM563407	AGENCOURT
828	19	0.5	896	13	BM018704	BM018704	603646589	901	19	0.5	1054	14	BM909670	BM909670	AGENCOURT
829	19	0.5	897	13	BM020378	BM020378	603649023	902	19	0.5	1055	14	BM810340	BM810340	AGENCOURT
830	19	0.5	899	12	BG677388	BG677388	602625207	903	19	0.5	1055	14	BM810583	BM810583	AGENCOURT
831	19	0.5	900	13	BM562944	BM562944	AGENCOURT	904	19	0.5	1055	14	BM909973	BM909973	AGENCOURT
832	19	0.5	901	14	BQ881261	BQ881261	AGENCOURT	905	19	0.5	1056	13	BM557786	BM557786	AGENCOURT
833	19	0.5	904	13	BM019746	BM019746	603648055	906	19	0.5	1056	14	BM811634	BM811634	AGENCOURT
834	19	0.5	905	17	CNS06002	AL408312	T7 end of	907	19	0.5	1058	13	BM562868	BM562868	AGENCOURT
835	19	0.5	913	13	BM017595	BM017595	603644784	908	19	0.5	1061	13	BM018263	BM018263	603645913
836	19	0.5	915	13	BM019763	BM019763	603648079	909	19	0.5	1061	13	BM551851	BM551851	AGENCOURT
837	19	0.5	916	12	BF694533	BF694533	602081745	910	19	0.5	1064	13	BM558175	BM558175	AGENCOURT
838	19	0.5	917	13	BI117140	BI117140	AGENCOURT	911	19	0.5	1065	13	BM045095	BM045095	603622966
839	19	0.5	922	14	BQ228619	BQ228619	AGENCOURT	912	19	0.5	1065	14	BM910941	BM910941	AGENCOURT
c 840	19	0.5	923	14	BQ879356	BQ879356	AGENCOURT	913	19	0.5	1066	14	BM909791	BM909791	AGENCOURT
841	19	0.5	929	12	BG677031	BG677031	602623677	914	19	0.5	1067	13	BM550164	BM550164	AGENCOURT
c 842	19	0.5	931	9	AL574617	AL574617	602623677	915	19	0.5	1068	13	BM550028	BM550028	AGENCOURT
843	19	0.5	933	12	BG675259	BG675259	602621548	916	19	0.5	1069	14	BM811546	BM811546	AGENCOURT
844	19	0.5	935	9	AL550868	AL550868	602621548	917	19	0.5	1070	14	BM811603	BM811603	AGENCOURT
845	19	0.5	940	12	BG698729	BG698729	602703110	918	19	0.5	1071	13	BM551550	BM551550	AGENCOURT
846	19	0.5	943	13	BM044795	BM044795	603622565	919	19	0.5	1072	13	BM562905	BM562905	AGENCOURT
c 847	19	0.5	945	12	BF788724	BF788724	602107124	920	19	0.5	1072	14	BM811582	BM811582	AGENCOURT
848	19	0.5	945	17	CNS03KHI	AL248175	Tetraodon	921	19	0.5	1075	14	BM910337	BM910337	AGENCOURT
849	19	0.5	948	12	BG678882	BG678882	602624785	922	19	0.5	1077	13	BM563716	BM563716	AGENCOURT
850	19	0.5	948	13	BI247034	BI247034	602960362	923	19	0.5	1078	14	BM910415	BM910415	AGENCOURT
851	19	0.5	955	17	CNS02A6K	AL188165	Tetraodon	924	19	0.5	1080	13	BM551640	BM551640	AGENCOURT
852	19	0.5	962	13	BM019087	BM019087	603647123	925	19	0.5	1080	14	BM909792	BM909792	AGENCOURT
853	19	0.5	966	12	BG677906	BG677906	602625605	926	19	0.5	1081	14	BM811582	BM811582	AGENCOURT
854	19	0.5	967	12	BG676549	BG676549	602623020	927	19	0.5	1082	14	BM908492	BM908492	AGENCOURT
855	19	0.5	969	14	BQ722262	BQ722262	AGENCOURT	928	19	0.5	1082	14	BM908895	BM908895	AGENCOURT
856	19	0.5	971	13	BM044841	BM044841	603622631	929	19	0.5	1083	13	BM563753	BM563753	AGENCOURT
857	19	0.5	974	13	BM020504	BM020504	603649222	930	19	0.5	1086	14	BM910407	BM910407	AGENCOURT
858	19	0.5	974	13	BM552300	BM552300	AGENCOURT	931	19	0.5	1087	14	BM811497	BM811497	AGENCOURT
859	19	0.5	976	13	BI084761	BI084761	602869781	932	19	0.5	1089	13	BM044878	BM044878	603622672
860	19	0.5	980	13	BM019512	BM019512	603647714	933	19	0.5	1095	14	BM910995	BM910995	AGENCOURT
861	19	0.5	983	13	BM563572	BM563572	AGENCOURT	934	19	0.5	1096	13	BM463971	BM463971	AGENCOURT
862	19	0.5	985	13	BM549782	BM549782	AGENCOURT	935	19	0.5	1096	14	BM910885	BM910885	AGENCOURT
863	19	0.5	988	13	BM563488	BM563488	AGENCOURT	936	19	0.5	1098	14	BM908597	BM908597	AGENCOURT
c 864	19	0.5	989	17	CNS05EAY	AL333475	Tetraodon	937	19	0.5	1100	17	CNS053Y1	AL321985	Tetraodon
865	19	0.5	990	13	BI649915	BI649915	603296560	938	19	0.5	1101	17	CNS053Y1	AL320050	Tetraodon
c 866	19	0.5	999	17	CNS03KRC	AL248277	Tetraodon	939	19	0.5	1103	13	BM551924	BM551924	AGENCOURT
867	19	0.5	1004	13	BM558067	BM558067	AGENCOURT	940	19	0.5	1107	13	BM551888	BM551888	AGENCOURT
868	19	0.5	1005	14	BM911079	BM911079	AGENCOURT	941	19	0.5	1117	13	BM551911	BM551911	AGENCOURT
869	19	0.5	1008	13	BM563148	BM563148	AGENCOURT	942	19	0.5	1121	13	BM019549	BM019549	603647766
870	19	0.5	1011	13	BI085197	BI085197	602870593	943	19	0.5	1121	14	BM914137	BM914137	AGENCOURT
c 871	19	0.5	1014	14	BM905487	BM905487	AGENCOURT	944	19	0.5	1127	13	BM549580	BM549580	AGENCOURT
872	19	0.5	1015	13	BM045460	BM045460	603623580	945	19	0.5	1128	14	BM909889	BM909889	AGENCOURT
873	19	0.5	1019	13	BM563658	BM563658	AGENCOURT	946	19	0.5	1133	14	BM911175	BM911175	AGENCOURT
874	19	0.5	1020	13	BM549875	BM549875	AGENCOURT	947	19	0.5	1138	14	BM909759	BM909759	AGENCOURT
875	19	0.5	1024	13	BM563777	BM563777	AGENCOURT	948	19	0.5	1143	14	BM913798	BM913798	AGENCOURT
c 876	19	0.5	1025	17	CNS01PNU	AL155050	Anopheles	949	19	0.5	1148	13	BI085564	BI085564	602870292
877	19	0.5	1025	13	BM550282	BM550282	AGENCOURT	950	19	0.5	1151	13	BM018036	BM018036	603645584
878	19	0.5	1028	13	BM563391	BM563391	AGENCOURT	951	19	0.5	1156	13	BM560737	BM560737	AGENCOURT
879	19	0.5	1030	13	BM549860	BM549860	AGENCOURT	952	19	0.5	1161	14	BM910187	BM910187	AGENCOURT
880	19	0.5	1031	12	BE799714	BE799714	601590665	953	19	0.5	1164	14	BM910908	BM910908	AGENCOURT
881	19	0.5	1031	13	BM549662	BM549662	AGENCOURT	954	19	0.5	1167	13	BM018893	BM018893	603646824
882	19	0.5	1031	13	BM552438	BM552438	AGENCOURT	955	19	0.5	1167	14	BM909105	BM909105	AGENCOURT

```
956 19 0.5 1171 13 BI085315 602870373
957 19 0.5 1174 13 BM549134 AGENCOURT
958 19 0.5 1176 13 BM045270 603623233
959 19 0.5 1176 14 BM911261 AGENCOURT
960 19 0.5 1180 13 BI085106 602870647
961 19 0.5 1190 13 BM549130 AGENCOURT
962 19 0.5 1198 14 BM908893 AGENCOURT
963 19 0.5 1204 13 BM549251 AGENCOURT
964 19 0.5 1204 14 BM914218 AGENCOURT
965 19 0.5 1205 13 BM017689 603645028
966 19 0.5 1207 14 BM910307 AGENCOURT
967 19 0.5 1208 14 BM908519 AGENCOURT
968 19 0.5 1213 14 BM909247 AGENCOURT
969 19 0.5 1222 14 BM908269 AGENCOURT
970 19 0.5 1222 14 BM910574 AGENCOURT
971 19 0.5 1224 14 BM908972 AGENCOURT
972 19 0.5 1234 14 BM911085 AGENCOURT
973 19 0.5 1235 13 BI085338 602870004
974 19 0.5 1242 14 BM913677 AGENCOURT
975 19 0.5 1243 13 BM018595 603646459
976 19 0.5 1243 14 BM911264 AGENCOURT
977 19 0.5 1248 13 BM017879 603645358
978 19 0.5 1251 13 BM045264 603623223
979 19 0.5 1251 14 BM911250 AGENCOURT
980 19 0.5 1252 14 BM908638 AGENCOURT
981 19 0.5 1256 13 BM045182 603623103
982 19 0.5 1260 13 BM549492 AGENCOURT
983 19 0.5 1266 14 BM913626 AGENCOURT
984 19 0.5 1271 13 BM018130 603645718
985 19 0.5 1281 14 BM910859 AGENCOURT
986 19 0.5 1287 14 BM908666 AGENCOURT
987 19 0.5 1289 14 BM910149 AGENCOURT
988 19 0.5 1298 14 BM908291 AGENCOURT
989 19 0.5 1309 14 BM911265 AGENCOURT
990 19 0.5 1316 14 BM913704 AGENCOURT
991 19 0.5 1321 14 BM914078 AGENCOURT
992 19 0.5 1341 13 BI081626 602880271
993 19 0.5 1342 14 BM910331 AGENCOURT
994 19 0.5 1343 14 BM909251 AGENCOURT
995 19 0.5 1351 14 BM908633 AGENCOURT
996 19 0.5 1359 13 BM557819 AGENCOURT
997 19 0.5 1362 14 BM910954 AGENCOURT
998 19 0.5 1377 14 BM909771 AGENCOURT
999 19 0.5 1377 14 BM908569 AGENCOURT
1000 19 0.5 1384 14 BM910087 AGENCOURT
```

ALIGNMENTS

```
RESULT 1
BF126187 553 bp mRNA linear EST 24-OCT-2000
LOCUS 601650450F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934045 5',
DEFINITION mRNA sequence.
ACCESSION BF126187
VERSION BF126187.1 GI:10965227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 553)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
```

```
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM774 row: d column: 14
High quality sequence stop: 542.
FEATURES
    source
        1..553
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3934045"
                /clone_lib="NIH_MGC_76"
                /lab_host="DH10B (TI phage-resistant)"
                /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                SfiI (ggccattatgcc); Site_2: SfiI (ggccattatgcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCGGAGGGCGGCAGATG-dt(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.85
                kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH_MGC Library."
    BASE COUNT 143 a 138 c 153 g 119 t
    ORIGIN
        Query Match 12.1%; Score 478; DB 12; Length 553;
        Best Local Similarity 100.0%; Pred. No. 1.9e-237;
        Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Qy 3486 TGTAGTACCCAAAGAGGTGAAGAACACCTGTGATCCTGTGAGAAACTACAGAACTGCGG 3545
    Db 1 TGTAGTACCCAAAGAGGTGAAGAACACCTGTGATCCTGTGAGAAACTACAGAACTGCGG 60
    Qy 3546 TGTTCGCCAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCGCCACCTGT 3605
    Db 61 TGTTCGCCAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCGCCACCTGT 120
    Qy 3606 GTTCTACTTGGAGAAAGACAATGAGACCATCCCTTGCACACAGAGAGGATCAGTATGCA 3665
    Db 121 GTTCTACTTGGAGAAAGACAATGAGACCATCCCTTGCACACAGAGAGGATCAGTATGCA 180
    Qy 3666 CCAGGACACACAGGGTATGCTGCTTCCTTCATTTCAGCCAGCCAGCAAGAAATCAGACGCTGG 3725
    Db 181 CCAGGACACACAGGGTATGCTGCTTCCTTCATTTCAGCCAGCCAGCAAGAAATCAGACGCTGG 240
    Qy 3726 ATGTTACACGTTGTTCAGCCAAAGATGAGCGCGATCGTGTGCTGCACTGCCAGCTGGA 3785
    Db 241 ATGTTACACGTTGTTCAGCCAAAGATGAGCGCGATCGTGTGCTGCACTGCCAGCTGGA 300
    Qy 3786 TATATACGCTCAGTGGCACCACATCAGATCCCAACGCCCATGTCTGTCCGGCCAGTGGCAG 3845
    Db 301 TATATACGCTCAGTGGCACCACATCAGATCCCAACGCCCATGTCTGTCCGGCCAGTGGCAG 360
    Qy 3846 TCGTACGGATCTCTCACCAGTAAGGACTTGCATATTTTCTGCCTTTTCTCCATGGA 3905
    Db 361 TCGTACGGATCTCTCACCAGTAAGGACTTGCATATTTTCTGCCTTTTCTCCATGGA 420
    Qy 3906 AAGCACCATGCTGTATTCATCTCTTCGAGTGTACTGAGAGTGTGATGAACATTAA 3963
    Db 421 AAGCACCATGCTGTATTCATCTCTTCGAGTGTACTGAGAGTGTGATGAACATTAA 478
RESULT 2
BF0218668
LOCUS AGENCOURT_7263964 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6017402
DEFINITION 5', mRNA sequence.
ACCESSION BF0218668
VERSION BF0218668.1 GI:20400068
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE 1 (bases 1 to 931)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13216 row: k column: 03
High quality sequence stop: 501.
FEATURES
source
1. .931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6017402"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."
BASE COUNT 250 a 225 c 236 g 220 t
ORIGIN
Query Match 7.3%; Score 291; DB 14; Length 931;
Best Local Similarity 100.0%; Pred. No. 8.4e-140;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3673 ACAACAGGGTATGCCCTTCATTCAGCCAGCCAGCAAGAAATCAGACGCTGGATGGTAC 3732
Db 1 ACAACAGGGTATGCCCTTCATTCAGCCAGCCAGCAAGAAATCAGACGCTGGATGGTAC 60
QY 3733 ACCTTGTCAGCCAAAGTGAAGCGGCATCGTGTGTCGACTGCCAGCGTGGATATATAC 3792
Db 61 ACCTTGTCAGCCAAAGTGAAGCGGCATCGTGTGTCGACTGCCAGCGTGGATATATAC 120
QY 3793 GCTCAGTGGCACCATCAGATGCCACCGCCATCTGTCTCGGCGCCAGTGGCAGTCGGTAC 3852
Db 121 GCTCAGTGGCACCATCAGATGCCACCGCCATCTGTCTCGGCGCCAGTGGCAGTCGGTAC 180
QY 3853 GGATCTCTACCAAGTAAAGGACTTGACATATTTTCTGCCCTTTCTCCATGGAAAGCAGC 3912
Db 181 GGATCTCTACCAAGTAAAGGACTTGACATATTTTCTGCCCTTTCTCCATGGAAAGCAGC 240
QY 3913 ATGGTGATTCATGCTCTCTCGAGTGTAGTGAGAGTGAATCAACTTTAA 3963
Db 241 ATGGTGATTCATGCTCTCTCGAGTGTAGTGAGAGTGAATCAACTTTAA 291
RESULT 3
LOCUS AO621045 503 bp DNA linear GSS 16-JUN-1999
DEFINITION HS_2221_B1_B06_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2221 Col=11 Row=D, DNA sequence.
ACCESSION AO621045
VERSION AO621045.1 GI:5083437
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 503)
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end web server: <http://www.htsc.washington.edu>
Plate: 2221 row: D column: 11
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 503.
FEATURES
source
1. .503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2221 Col=11 Row=D"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 154 a 122 c 106 g 115 t 6 others
ORIGIN
Query Match 6.8%; Score 270; DB 17; Length 503;
Best Local Similarity 100.0%; Pred. No. 6.8e-129;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 68 AAACCCAGACATCGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAAACCTTGCC 127
Db 93 AAACCCAGACATCGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAAACCTTGCC 152
QY 128 ATTTGCGCAGTCTCTCTGCGGCGCTGAAGAGGCGGAGGCGCAAGATGACCTTCCAGATC 187
Db 153 ATTTGCGCAGTCTCTCTGCGGCGCTGAAGAGGCGGAGGCGCAAGATGACCTTCCAGATC 212
QY 188 TTTTCAGCCTTTTCAGCCCAAGAAGAAATAGACGAAAGTGTCAATTTGGCAGACTGGCCA 247
Db 213 TTTTCAGCCTTTTCAGCCCAAGAAGAAATAGACGAAAGTGTCAATTTGGCAGACTGGCCA 272
QY 248 TCAATTACGACCCCTTTGGAGAAGGCAGATGAACCTCAAGCTAGAAAACGACTTTCTCTG 307
Db 273 TCAATTACGACCCCTTTGGAGAAGGCAGATGAACCTCAAGCTAGAAAACGACTTTCTCTG 332
QY 308 ATCAGATGAACACTCACCCTAAATTTAAAGT 337
Db 333 ATCAGATGAACACTCACCCTAAATTTAAAGT 362
RESULT 4
LOCUS AA179599 481 bp mRNA linear EST 31-DEC-1996
DEFINITION Zp49f09.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:612809 5' similar to TF:G1160355 G1160355 UNC-89.; mRNA sequence.
ACCESSION AA179599
VERSION AA179599.1 GI:1760985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 481)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, J.B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisakis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.
TITLE Generation and analysis of 280,000 human expressed sequence tags

1026


```
JOURNAL      Genome Res. 6 (9), 807-828 (1996)
MEDLINE      97044478
COMMENT      Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LLNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: -28kl3 rev2 from Amersham
              High quality sequence stop: 417.
              Location/Qualifiers
                1. .481
                /organism="Homo sapiens"
                /db_xref="GDB:464648"
                /db_xref="taxon:9606"
                /clone="IMAGE:612809"
                /clone_lib="Stratagene HeLa cell s3 937216"
                /sex="female"
                /dev_stage="HeLa S3 cell line"
                /lab_host="SOLR (kanamycin resistant)"
                /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
              xhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3
              epithelioid carcinoma cells grown to semi-confluency
              without induction. Average insert size: 1.5 kb; Uni-ZAP XR
              Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
              adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'
              2 others
              BASE COUNT      129 a 132 c 125 g 93 t
              ORIGIN
              Query Match      5.3%; Score 212; DB 9; Length 481;
              Best Local Similarity 99.7%; Pred. No. 1.2e-98;
              Matches 332; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
              QY 3423 AGGACCTATAAGTGCATCGCTACCAACAAACCGGCGAGAAATCTTTTAGTCFGGAGCT 3482
              Db 148 AGGACCTATAAGTGCATCGCTACCAACAAACCGGCGAGAAATCTTTTAGTCFGGAGCT 207
              QY 3483 CTCGTGTAGTACCAAGAGGTGAAGAAGCACCTGTGATCTCGAGAACTACAGAACTG 3542
              Db 208 CTCGTGTAGTACCAAGAGGTGAAGAAGCACCTGTGATCTCGAGAACTACAGAACTG 267
              QY 3543 CGGTGTTCCGAAAGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACCC 3602
              Db 268 CGGTGTTCCGAA -GCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACCC 326
              QY 3603 TGTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGACACGAGAGGATCAGTAT 3682
              Db 327 TGTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGACACGAGAGGATCAGTAT 386
              QY 3663 GCACGAGGACACACAGGGTATGCTGCTTCTCATTTCAGGCAGCCAAAGAAATCAGACGC 3722
              Db 387 GCACGAGGACACACAGGGTATGCTGCTTCTCATTTCAGGCAGCCAAAGAAATCAGACGC 446
              QY 3723 TGGATGTACACGTTGTGACGCCAAGATGAAGC 3755
              Db 447 TGGATGTACACGTTGTGACGCCAAGATGAAGC 479
              RESULT 5
              AQ167332/c
              LOCUS
              DEFINITION
              HS_3173_A1_A03_MR CIT Approved Human Genomic Sperm Library D Homo
              sapiens genomic clone Plate-3173 Col-5 Row-A, DNA sequence.
              ACCESSION
              AQ167332
              VERSION
              AQ167332.1 GI:3565497
              KEYWORDS
              GSS.
              SOURCE
              human.
              ORGANISM
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 331)
              REFERENCE
              1
              Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
              Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
              Hood, L.
              Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
              Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
              99380589
              COMMENT
              Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Sequence Tagged Connector
              Plate: 3173 row: A column: 5
              Class: BAC ends
              High quality sequence stop: 331.
              Location/Qualifiers
                1. .331
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate-3173 Col-5 Row-A"
                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
              E-Coli DH10B"
                BASE COUNT      81 a 77 c 86 g 87 t
                ORIGIN
                Query Match      4.2%; Score 165; DB 17; Length 331;
                Best Local Similarity 100.0%; Pred. No. 3.9e-74;
                Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
                QY 3297 ACCGCCCGGAGTGACATGGCTACTCAATGGCAACCTGTGTACAGATGCTCCCA 3356
                Db 165 ACCGCCCGGAGTGACATGGCTACTCAATGGCAACCTGTGTACAGATGCTCCCA 106
                QY 3357 CAACATGCTGTGAGGAGACCGGAGTCCACTCTCTCTCATTGACCCACTCAGCG 3416
                Db 105 CAACATGCTGTGAGGAGACCGGAGTCCACTCTCTCTCATTGACCCACTCAGCG 46
                QY 3417 CGACGCGAGGACCTTAAAGTGCATCGCTACCAACAAACCGGCA 3461
                Db 45 CGACGCGAGGACCTTAAAGTGCATCGCTACCAACAAACCGGCA 1
              RESULT 6
              AQ768080
              LOCUS
              DEFINITION
              HS_3097_B1_E02_MR CIT Approved Human Genomic Sperm Library D Homo
              sapiens genomic clone Plate-3097 Col-3 Row-J, DNA sequence.
              ACCESSION
              AQ768080
              VERSION
              AQ768080.1 GI:5646196
              KEYWORDS
              GSS.
              SOURCE
              human.
              ORGANISM
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 439)
              REFERENCE
              1
              Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
              Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
              Hood, L.
              Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
              Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
              99380589
              COMMENT
              Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
```


Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3097 row: J column: 3
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 439.
Location/Qualifiers
1. .439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Place=3097 Col=3 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

FEATURES

source

BASE COUNT 105 a 127 c 89 g 117 t 1 others
ORIGIN

Query Match 4.0%; Score 159; DB 17; Length 439;
Best Local Similarity 99.5%; Pred. No. 5.6e-71;
Matches 209; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3284 AGGTGAGTGTTCACGCCGCCGCCGAGCTGACATGGCTACTCAATGGCCAACTGTGTCTAC 3343
|||||
Db 67 AGGTGAGTGTTCACGCCGCCGCCGAGCTGACATGGCTACTCAATGGCCAACTGTGTCTAC 126
|||||

QY 3344 CAGATCGCTCCACAGATCGTGTGAGGAGACCGGAGTCCACTCTCTCTCATTTGACC 3403
|||||
Db 127 CAGATCGCTCCACAGATCGTGTGAGGAGACCGGAGTCCACTCTCTCTCATTTGACC 186
|||||

QY 3404 CACTCACTACGCGCAGCGAGGACCTATTAAGTGCATCGTACCAACAAACCGGGCAGA 3463
Db 187 CACTCACTACGCGCAGCGAGGACCTATTAAGTGCATCGTACCAACAAACCGGGCAGA 246
|||||

QY 3464 ATTCCTTTTACTGTGGAGCTCTCTGTAGTAG 3493
|||||
Db 247 ATTCCTTTTACTGTGGAGCTCTCTGTAGTAG 276
|||||

RESULT 7

BM968035

LOCUS

DEFINITION LM24HW00787 Bos taurus LM-24-HW cdna library Bos taurus cDNA clone
LM-24-HW-007-87 (5'), mRNA sequence.

ACCESSION

BM968035

VERSION

BM968035.1

KEYWORDS

EST.

SOURCE

cow.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 614)

Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,

Sun,S.S. and Cheong,I.C.

Gene Expression Profiling of the Bovine skeletal muscle

Unpublished (2002)

Contact: Dr. Du-Hak Yoon

National Livestock Research Institute, RDA

564 Omockchun-dong, Suwon, 441-350, Korea

Tel: 82 31 290 1593

Fax: 82 31 290 1792

Email: dhyoon@rda.go.kr

Insert Length: 614 Std Error: 0.00

Seq primer: CAGGAACAGCTATGAC

POLYA=No.

Location/Qualifiers

1. .614

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone="LM-24-HW-007-87 (5')"

/clone_lib="Bos taurus LM-24-HW cdna library"

FEATURES

source

/sex="six males mixed"
/tissue_type="longissimus dorsi"
/cell_type="myocyte"
/dev_stage="24 months old"
/lab_host="XLI-BlueMRF strain"
/note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site_1:
EcoRI; Site_2: Xho I"
BASE COUNT 150 a 188 c 159 g 117 t
ORIGIN

Query Match 1.4%; Score 55; DB 14; Length 614;
Best Local Similarity 100.0%; Pred. No. 1.1e-16;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 CCTCCCAAGATGCTGTCAGGAGACCGGAGTCCACTCTCTCATTTGACCC 3404
|||||
Db 144 CTTCCCAAGATGCTGTCAGGAGACCGGAGTCCACTCTCTCATTTGACCC 198
|||||

RESULT 8
AG083688/c

LOCUS AG083688 633 bp DNA linear GSS 03-NOV-2001

DEFINITION Pan troglodytes DNA, clone: PTB-081D08.R, genomic survey sequence.

ACCESSION AG083688

VERSION AG083688.1

KEYWORDS GSS.

SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male

BAC Library clone:PTB-081D08.R.

ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 633)

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-Chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chuhpbes@sc.riken.go.jp, URL:http://hgpc.gsc.riken.go.jp/.

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI.

1. .633

Location/Qualifiers

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-081D08.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

169 a 140 c 109 g 213 t 2 others

BASE COUNT

ORIGIN

Query Match 1.2%; Score 48; DB 17; Length 633;

Best Local Similarity 100.0%; Pred. No. 5.1e-13;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1077 AGGGTTTCTTCTTCTGACTCAGAGCGGACCCCTAACGAAGAGAT 1124

|||||

Db 375 AGGGTTTCTTCTTCTGACTCAGAGCGGACCCCTAACGAAGAGAT 328

|||||

```

RESULT 9
AW918754
LOCUS EST350058 Rat gene index, normalized rat, norvegicus, Banto Soares
DEFINITION Rattus norvegicus cDNA clone RGIEX58 5' end, mRNA sequence.
ACCESSION AW918754
VERSION AW918754.1 GI:8084539
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 376)
AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.
Location/Qualifiers
source 1. .376
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RGIEX58"
/clone_lib="Rat gene index, normalized rat, norvegicus,
Bento Soares"
/tissue_type="mix - brain, ovary, placenta, kidney, lung,
liver, embryo, heart, muscle, spleen"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Estimated insert size approx.1 kb"
BASE COUNT 107 a 96 c 97 g 76 t
ORIGIN
Query Match 1.0%; Score 41; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2890 TTCACCTGCAAAATGTTGGGATACCTGTTCCAAAGGTTTA 2930
|||||
Db 182 TTCACCTGCAAAATGTTGGGATACCTGTTCCAAAGGTTTA 222

RESULT 10
BQ559749
LOCUS H4060B02-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H4060B02 5', mRNA sequence.
ACCESSION BQ559749
VERSION BQ559749.1 GI:21460634
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 413)
AUTHORS VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin
, P.R., Stagg, C.A., Bassey, U., Alba, K., Hamatani, T., Kargul, G.J.,
Luo, A.G. and KO, M.S.H.
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set
JOURNAL Unpublished (2002)
COMMENT Other_ESTs: H4060B02-3

```

```

Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cdna/NIA\_7.4k.html for details.
Plate: H4060 row: B column: 02
Seq primer: -21M13 Reverse
High quality sequence stop: 413
POLYA-No.
Location/Qualifiers
source 1. .413
/organism="Mus musculus"
/db_xref="C57BL/6"
/db_xref="niaEST:H4060B02-5"
/db_xref="taxon:10090"
/clone="H4060B02"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."
BASE COUNT 102 a 123 c 114 g 74 t
ORIGIN
Query Match 1.0%; Score 38; DB 14; Length 413;
Best Local Similarity 100.0%; Pred. No. 7.7e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3196 CTACAGGAACGCTTTTCCGACACACATTTCTCGAGGC 3233
|||||
Db 73 CTACAGGAACGCTTTTCCGACACACATTTCTCGAGGC 110

RESULT 11
AZ889597/c
LOCUS RPCI-24-167B6.TJ RPCI-24 Mus musculus genomic clone RPCI-24-167B6,
DEFINITION DNA sequence.
ACCESSION AZ889597
VERSION AZ889597.1 GI:13208542
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 503)
AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akkinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-167B6-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 167 row: B column: 6
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
source 1. .503

```

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shinbata, K. and Hayashizaki, Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken, Genomic Sciences Center and Genome Science Laboratory in Riken, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.									
FEATURES	Location/Qualifiers								
source	1. 513								
	/organism="Mus musculus"								
	/db_xref="taxon:10090"								
	/clone="9830112703"								
	/clone_lib="RIKEN full-length enriched, adult male bone"								
	/sex="male"								
	/tissue_type="bone"								
	/dev_stage="adult"								
	/lab_host="DH108"								
	/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken, Genomic Sciences Center and Genome Science Laboratory in Riken, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."								
BASE COUNT	129	a	166	c	113	g	105	t	
ORIGIN									
	Query Match 0.8%; Score 32; DB 10; Length 513;								
	Best Local Similarity 100.0%; Pred. No. 0.00011;								
	Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
QY	1978	TCCACTCAGTTACAAACAGCTTCATAACCAAGT 2009							
Db	223	TCCACTCAGTTACAAACAGCTTCATAACCAAGT 254							
RESULT 13									
BH261787/c									
LOCUS	BH261787 521 bp DNA linear GSS 30-NOV-2001								
DEFINITION	CH230-90K20-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone								
	CH230-90K20, DNA sequence.								
ACCESSION	BH261787								
VERSION	BH261787.1								
KEYWORDS	GI:17167683								
SOURCE	GSS.								
ORGANISM	Norway rat.								
	Rattus norvegicus								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.								
REFERENCE	1 (bases 1 to 521)								
AUTHORS	Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., Jong, P. and Fraser, C.M.								
TITLE	Rat BAC End Sequences from Library CHORI-230 EcoRI segment								
JOURNAL	Unpublished (1999)								
COMMENT	Other_GSSs: CH230-90K20.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics								

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhac@tigr.org

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(<http://www.chori.org/bacpac/oreringinformation.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 90 row: K column: 20

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers
source
1..521
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-90K20"
/cell_type="Brain"
/sex="Female"
/note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT 145 a 117 c 117 g 142 t

ORIGIN

Query Match 0.8%; Score 31; DB 17; Length 521;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3737 TGTGAGCCAGCAATGAAGCGGCATCGTGC 3767

Db 282 TGTGAGCCAGCAATGAAGCGGCATCGTGC 252

RESULT 14
LOCUS AW315222 438 bp mRNA linear EST 25-APR-2001
DEFINITION 13252 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW315222
VERSION AW315222.1 GI:6744478
KEYWORDS EST.
SOURCE cow.

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
1 (bases 1 to 438)
Casas,E., Wray,J.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

COMMENT Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.

PCR Primers

FORWARD: AGCAACAGCTATGACCAT

BACKWARD: GTTTCCCGAGTCACGCG

Plate: 7 row: B column: 24

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

source
1..438

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 3BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;

Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal

longissimus muscle."

BASE COUNT 121 a 110 c 119 g 88 t

ORIGIN

Query Match 0.7%; Score 29; DB 10; Length 438;

Best Local Similarity 100.0%; Pred. No. 0.0039;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3553 GAAGGCCACCCGTGAGACTGGAGTGCGG 3581

Db 145 GAAGGCCACCCGTGAGACTGGAGTGCGG 173

RESULT 15

LOCUS BB633981

DEFINITION BB633981 RIKEN full-length enriched, adult male spinal cord Mus

musculus cDNA clone A330094N09 5', mRNA sequence.

ACCESSION BB633981

VERSION BB633981.1 GI:16470313

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 652)

REFERENCE Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda

, Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki

, D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resesgsc.riken.go.jp,

URL:<http://genome.gsc.riken.go.jp/>

, M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

, S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

, Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 19:02:20 ; Search time 36 seconds
(without alignments)
7555.061 Million cell updates/sec

Title: US-09-818-990B-2
Perfect score: 6890
Sequence: 1 MQDDIEASTISQLLRESY.....MESTMYSCSRVVSDEL 1320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriaph.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6869	99.7	1320	4 Q96KF5	Q96KF5 homo sapien
2	2560	37.2	507	4 Q96K90	Q96K90 homo sapien
3	2213.5	32.1	1106	4 Q8WX93	Q8WX93 homo sapien
4	1666	24.2	772	4 Q9Y2J6	Q9Y2J6 homo sapien
5	1460	21.2	492	11 Q9ET54	Q9ET54 mus musculus
6	1289.5	18.7	385	4 Q9UQF5	Q9UQF5 homo sapien
7	1289.5	18.7	404	4 Q9Y3E9	Q9Y3E9 homo sapien
8	1271.5	18.5	379	11 Q9CWM1	Q9CWM1 mus musculus
9	871.5	12.6	4162	13 Q98918	Q98918 gallus gall
10	826	12.0	2164	13 Q9IAR9	Q9IAR9 gallus gall
11	801.5	11.6	26926	4 Q10466	Q10466 homo sapien
12	801.5	11.6	26926	4 Q8WZB3	Q8WZB3 homo sapien
13	793	11.3	2000	6 Q97791	Q97791 oryctolagus
14	780.5	11.3	5604	4 Q8WZ53	Q8WZ53 homo sapien
15	780.5	11.3	34350	4 Q8WZ42	Q8WZ42 homo sapien
16	729	10.6	498	4 Q9UBF9	Q9UBF9 homo sapien

17	704	10.2	496	11 Q9JIF9	Q9JIF9 mus musculus
18	602.5	8.7	4824	5 Q95YM1	Q95YM1 procambarius
19	602.5	8.7	17352	5 Q95YM2	Q95YM2 procambarius
20	600	8.7	6632	5 Q01761	Q01761 caenorhabdi
21	598	8.7	6632	5 Q17362	Q17362 caenorhabdi
22	585.5	8.5	4816	5 Q8TI03	Q8TI03 bombyx mori
23	568.5	8.3	1561	11 Q924D2	Q924D2 mus musculus
24	545.5	7.9	4796	5 Q9W055	Q9W055 drosophila
25	545.5	7.9	16215	5 Q9NFS3	Q9NFS3 drosophila
26	542	7.9	4001	5 Q9N2P7	Q9N2P7 drosophila
27	542	7.9	4796	5 Q9NL88	Q9NL88 drosophila
28	539	7.8	1802	6 Q28633	Q28633 oryctolagus
29	536	7.8	2673	4 Q96SC3	Q96SC3 homo sapien
30	534	7.8	5636	4 Q96RW7	Q96RW7 homo sapien
31	531.5	7.7	7962	4 Q10465	Q10465 homo sapien
32	503	7.3	2154	4 Q8WZ51	Q8WZ51 homo sapien
33	502	7.3	2541	5 Q19663	Q19663 caenorhabdi
34	501.5	7.3	4203	5 Q965G2	Q965G2 caenorhabdi
35	501.5	7.3	4219	5 Q9NL87	Q9NL87 caenorhabdi
36	493.5	7.2	4488	5 Q9TXK2	Q9TXK2 caenorhabdi
37	485	7.0	6620	4 Q96AA2	Q96AA2 homo sapien
38	482	7.0	4650	4 Q15598	Q15598 homo sapien
39	472.5	6.9	3262	11 Q9EQJ5	Q9EQJ5 mus musculus
40	472	6.9	6815	5 Q9I7U4	Q9I7U4 drosophila
41	468.5	6.8	5198	5 Q76518	Q76518 caenorhabdi
42	467	6.8	1232	4 Q8TCG8	Q8TCG8 homo sapien
43	459.5	6.7	1019	4 Q9Y6L9	Q9Y6L9 homo sapien
44	453.5	6.6	1252	4 Q96DN3	Q96DN3 homo sapien
45	435.5	6.3	3354	5 Q8TI01	Q8TI01 bombyx mori

ALIGNMENTS

RESULT 1

Q96KF5 PRELIMINARY; PRT; 1320 AA.

AC Q96KF5; (TREMREL. 19, Created)
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
DE Myopalladin.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SKELETAL MUSCLE;

RX MEDLINE=21206024; PubMed=11309420;

RA Bang M.L., Mudry R.E., McElhinny A.S., Trombitas K., Geach A.J.,

Yamasaki R., Sorimachi H., Granzier H., Gregorio C.C., Labeit S.;

RT "Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple

roles in z-disc and I-band protein assemblies.";

RL J. Cell Biol. 153:413-428(2001).

DR EMBL: AF328296; AAK50625.1; -

DR InterPro: IPR000886; ER_target.

DR InterPro: IPR003006; Ig_MHC.

DR Pfam: PF00047; ig; 5.

DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.

SQ SEQUENCE 1320 AA; 145379 MW; A6579FB164D33B6E CRC64;

Query Match 99.7%; Score 6869; DB 4; Length 1320;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 1314; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MQDDIEASTISQLLRESYLAETRRHGNRSRAEPSSNCPHFSGPSGAAGGGGDDL 60

Db 1 MQDDIEASTISQLLRESYLAETRRHGNRSRAEPSSNCPHFSGPSGAAGGGGDDL 60

QY 61 PDL5AFLSQEELDSVNLRLAINYDPLEKADETQARKRLSPDQMKHSPNLSFEPNFCQD 120

Db 61 PDL5AFLSQEELDSVNLRLAINYDPLEKADETQARKRLSPDQMKHSPNLSFEPNFCQD 120

Function on hand
based on homology

Qy	121	NPRSPTSSKESPOEAKRPQYCS	STQSKVFLNKAADFTIELSSLPKSHSSKRIIPRACKN	180
Db	121	NPRSPTSSKESPOEAKRPQYCS	STQSKVFLNKAADFTIELSSLPKSHSSKRIIPRACKN	180
Qy	181	HKSLESQNKVMOENSSFDLSRERS	SSVPIPIADTRONEVNHAEQOEAKRREAQ	240
Db	181	HKSLESQNKVMOENSSFDLSRERS	SSVPIPIADTRONEVNHAEQOEAKRREAQ	240
Qy	241	AASEAAGDTPGSSPSSLYEELPQ	PPRFQKLRSREVPEGTRVQLDCITVVGIPPPQV	300
Db	241	AASEAAGDTPGSSPSSLYEELPQ	PPRFQKLRSREVPEGTRVQLDCITVVGIPPPQV	300
Qy	301	RWYCEGKELNSPDIHIVQAGNLH	SUTIAAEFEEDTGRYSCFASNIYGTDTSAEYIEG	360
Db	301	RWYCEGKELNSPDIHIVQAGNLH	SUTIAAEFEEDTGRYSCFASNIYGTDTSAEYIEG	360
Qy	361	VSSSDSGDPNKKEMNRIKPN	EVSSPPTTSAVIPPAVPOAQHLVAOPRVATIOQCOSP	420
Db	361	VSSSDSGDPNKKEMNRIKPN	EVSSPPTTSAVIPPAVPOAQHLVAOPRVATIOQCOSP	420
Qy	421	NYLQGLDGKPIIAAPFTKMLQ	LSASEQLVVFECRVKGAPSPKVEWYREGTLIEDSPD	480
Db	421	NYLQGLDGKPIIAAPFTKMLQ	LSASEQLVVFECRVKGAPSPKVEWYREGTLIEDSPD	480
Qy	481	FRILOKPRSMAPBEICTLVIAE	VAEDSGGCTCTASNKYGTVSSIAQLHVRGNEOLSN	540
Db	481	FRILOKPRSMAPBEICTLVIAE	VAEDSGGCTCTASNKYGTVSSIAQLHVRGNEOLSN	540
Qy	541	NGSLHSANSTTNLAAETOP	SPPHSEPPSVQOPPKPLEGVLVNHNEPRSSSRIGLRAVHF	600
Db	541	NGSLHSANSTTNLAAETOP	SPPHSEPPSVQOPPKPLEGVLVNHNEPRSSSRIGLRAVHF	600
Qy	601	NLPEDDKGSEASSEAGVVT	TROTROPDSQERENGQATKTPEPSPKPPVLAAPKLDS	660
Db	601	NLPEDDKGSEASSEAGVVT	TROTROPDSQERENGQATKTPEPSPKPPVLAAPKLDS	660
Qy	661	TQLOQLHNOVLLBQHQLO	NPSPSPKPEFPFXMTVLNSNAPPAVTTSKQVKAPSSQTFSL	720
Db	661	TQLOQLHNOVLLBQHQLO	NPSPSPKPEFPFXMTVLNSNAPPAVTTSKQVKAPSSQTFSL	720
Qy	721	ARPKYFPSTNTTAATVAPSS	PVFTLSSTPQTIQRTVSKESLVLVSHSPSVQTKSPGGLSI	780
Db	721	ARPKYFPSTNTTAATVAPSS	PVFTLSSTPQTIQRTVSKESLVLVSHSPSVQTKSPGGLSI	780
Qy	781	QNEPLPGPTEPTPPPTT	SIPSGNQFQPRCVSPIPVSPTSRIQNPVAFVSSVLPSLPAI	840
Db	781	QNEPLPGPTEPTPPPTT	SIPSGNQFQPRCVSPIPVSPTSRIQNPVAFVSSVLPSLPAI	840
Qy	841	PPTNAMGLPNSAPSMPSQ	LAKNTKSPQVNDNIRETKNAVITDGLCKKTTFSDVRPNQ	900
Db	841	PPTNAMGLPNSAPSMPSQ	LAKNTKSPQVNDNIRETKNAVITDGLCKKTTFSDVRPNQ	900
Qy	901	QEKYKISSFEORLNEIF	RLERTPVDESDEIQHDEIPTGKCIAPIDFKRLKHFRVTEGS	960
Db	901	QEKYKISSFEORLNEIF	RLERTPVDESDEIQHDEIPTGKCIAPIDFKRLKHFRVTEGS	960
Qy	961	PVTCTKIVGIPVKYVWF	KDGKQISKRNEHCKMRREGDGTCSUHIESTTSDDDGNTYIM	1020
Db	961	PVTCTKIVGIPVKYVWF	KDGKQISKRNEHCKMRREGDGTCSUHIESTTSDDDGNTYIM	1020
Qy	1021	AANPQGRISCGHLMVQSL	PIRSRLTSAGOSHRGSRVQERDKPELOERFPRPHFLOAPG	1080
Db	1021	AANPQGRISCGHLMVQSL	PIRSRLTSAGOSHRGSRVQERDKPELOERFPRPHFLOAPG	1080
Qy	1081	DMVAHEGRLCRLCKV	SGLPPPPPELTWLLNGQPVLPDASHKMLVRETGVHSLLDPLTQRD	1140
Db	1081	DMVAHEGRLCRLCKV	SGLPPPPPELTWLLNGQPVLPDASHKMLVRETGVHSLLDPLTQRD	1140
Qy	1141	AGTYKCIATNKTQNS	FSLELSVVAKEVKKAPILEIKLQNCVGPVGHVRLCEKRVGMPP	1200
Db	1141	AGTYKCIATNKTQNS	FSLELSVVAKEVKKAPILEIKLQNCVGPVGHVRLCEKRVGMPP	1200

Qy	1201	PVFYWKDNETIPTCRRRISNMHODTGYACLLIQPAKKSADAGWYLSAKNEAGIVSCTAR	1260
Db	1201	PVFYWKDNETIPTCRRRISNMHODTGYACLLIQPAKKSADAGWYLSAKNEAGIVSCTAR	1260
Qy	1261	LDIYAQHHQIIPPPMSVRPSGSRYSGLTSKGLDIFSAFSSMESTMWYSCSSRSVWSEDEL	1320
Db	1261	LDIYAQHHQIIPPPMSVRPSGSRYSGLTSKGLDIFSAFSSMESTMWYSCSSRSVWSEDEL	1320
RESULT 2			
Q956K90			
ID	Q956K90	PRELIMINARY;	PRT; 507 AA.
AC	Q956K90;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	CDNA FLJ11437 fis, clone HEMBB1000083, weakly similar to myosin light		
DE	chain kinase, smooth muscle and NON-muscle isozymes		
DE	(EC 2.7.1.117).		
OS	Homo sapiens (Human).		
OC	Eumaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N. A.		
RC	TISSUE=EMBRYO;		
RA	Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yanamoto J., Wakamatsu K., Nakamura Y., Nagahara K., Masuho Y.,		
RA	Ninomiya K., Iwayanagi T.		
RT	"NDO human cDNA sequencing project."		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK027343; BAB55048.1; -		
DR	InterPro; IPR003006; Ig_MHC.		
DR	Pfam; PF00047; Ig_2.		
SO	SEQUENCE 507 AA; 56185 MW; BD8488478845DF39 CRC64;		


```
|||||
Db 421 NYLQGLDGRPIIAAPVFTKMLQNSAQGLVVFECRVKAGSPKVEWYREGTLIEDSPD 480
QY 481 FRILQKKPRSMARP 494
Db 481 FRILQKKPRSMARP 494
|||||

RESULT 3
Q8WX93 PRELIMINARY; PRT; 1106 AA.
AC Q8WX93;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Myoneurin.
GN MYN.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lockwood S.K.; Sims K.B.;
RT "A protein with two immunoglobulin-like domains interacts with the
RT Norrie disease gene product norrin."
RL Am. J. Hum. Genet. 61:A158-A158(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Lockwood S.K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF464873; AAL69964.1; -.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000634; S/T_dehydrtse.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00409; Ig; 5.
DR SMART: SM00408; IgC2; 5.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; UNKNOWN_1.
SQ SEQUENCE 1106 AA; 122019 MW; 80B52268DD90E7C6 CRC64;

Query Match 32.1%; Score 2213.5; DB 4; Length 1106;
Best Local Similarity 38.1%; Pred. No. 3.1e-139;
Matches 518; Conservative 152; Mismatches 305; Indels 383; Gaps 26;

QY 38 SNNPCHFGSPGAEGGGGDDLPDLAFLSQEELDESVMRLAIINYPLEKADETOAR 97
Db 5 SSHEFYDLSDMQESKNTDFPGLSAFLSQEELDESVMRLAIINYPLEKADETOAR 64
QY 98 KRL---SPDQMKHP-----NLSEFNFQDNPSP-----TSSKESPOEAK- 136
Db 65 SQIFSTSPALCEHPSHKETHKEGASRRP---QDNSTVPQVLAETKISSPVSKRK 121
QY 137 -----RPQYC-----SETQSKV-----FLNKAADFIEELSS 163
Db 122 PAMSPLLTRPSYIRSLRAEKRAKTPSTNVKPTPHQRKGQSQQLCDKRAANLIELTS 181
QY 164 LFYSKSHSKRTRPRACKNHNKSLKESQNVQMSNSSFSDLSERRERSVPPIPADTRNE 223
Db 182 IFKAAPRNPNRPNNG-----ESSPDSGYLSKPNQPSALISASAS----- 220
QY 224 VNHALEQEAQKRAEQAASAAAGDTPPGSSP-----SSLYEPLGPPPRFTQ 273
Db 221 -QSPMEDQGEWEREVKSPGARHCYQDQNDLAVPHNRKSHFQPHSALHFP-----AAPRFQ 275
QY 274 KLRSEVPEGTRVQDLCVIVGIPPPQVRWYCEGKEGKESNPDHIF-VQAGNLHSLTIAEAF 332
Db 276 KLRQEAEGSVRYLECRVTGNTPTRVRFCEGKELHNTPDQIHCEGDLHTLIIAEAF 335
QY 333 EDTGRYSCFASNIYGDSTSAEIIYIGVSSSSDEGD-----PNKEMNRIQRPNE----- 383
Db 336 EDDTGRTCLATNPSGSDTTSAEVFIEGASSTDSDESLEAFKSRAGAMPOAKKTTSVSL 395
```

```
QY 384 ---VSSPPT---TSVIPPVAPQAHVAQPRVATIQCCQSPNTYVLOGLDG-KPIIAAPVF 437
Db 396 TIGSSPKTGVTAVI-----QPLSVVQOVHSPSTYLCRDPDGTATYFPVF 443
QY 438 TKMLQNSAQGLVVFECRVKAGSPKVEWYREGTLIEDSPFRILQKKPRSMARPEEI 497
Db 444 TKELQNTAVAEQGVVLECRVGRAPPLQVQWFRQGSIQDSPFRILQKKPRSTAEPEEI 503
QY 498 CTLVIAEVAEDSGCTCTASNKYGTVSSIAQLHVRGNEDELNNGLSHSANSNTLNAAIE 557
Db 504 CTLVIAETFPEDAGIFTCSARNIDYGSATQAQLVVTGSAN--TENCSTYESGESNNDHFQH 561
QY 558 PQSPHSEPPSVEQPKPKLEGVLVNHNEPRSSRILGRVHFNLPDDDDKSGSEASSAGV 617
Db 562 FPPPPPILETSSLELASKPKPSEIQOVN-NPELGLSRAALQWQFNAAERE-----TSGV 613
QY 618 VTTRQTRPDSXQERFNGQATKT---PEPSF---PVKEPPPVLAAPKPKLDSTQLQOLHNOVL 671
Db 614 HPSR-----GVNGLINGKANSKSLPTPAVLLSPTKEPPPLAKPKL----- 655
QY 672 LEQHLQNPSPSPKPEFPFVMTVLNSNAPPVATTSXKQVKAPSSQTSFSLARPKYFFPSTN 731
Db 656 ----- 655
QY 732 TTAATVAPSSSPVFTLSSTPTQIQTIVTSKESLLVSHPSVQTKSPGLSIQNEPLPPGPT 791
Db 656 ----- 655
QY 792 PTPPPTFSIPSGNQFQPCVPIPVSPTRIQNPVAFLSVLPSLPAIPPTNAXMLPRS 851
Db 656 ----- 655
QY 852 APSMPSQGLAKNNTKSPQVNDONIRETKNAVIRDLGKKKITFSDVRPN----- 899
Db 656 -----GFPKASRTARIASDEIOGTQDAVTDLERKLRFKEDLLNNGQPRLTYEERM 708
QY 900 -----QOEYKISSFEORLANEIEFRLERTPVDESDDDEIQHDE 936
Db 709 ARLLGADSATVFNIOPEBETANQEVKVSCEQRLISEIYRLERSPVDESQVQYGD 768
QY 937 IPTGKCIAPIDKRLKHFRTVEGSPVFTCKIVGIPVYVFKDQKQISKRNEHCKMR 996
Db 769 VPVENGMAPFEFMKLKHKIFEGMPVFTCRVAGNPKPKIYWFKDGQISPKSDHYTIOR 828
QY 997 EGDGTCSLHIESITSDDDGNYTMAANPQGRISCSGHLVQSLPIRSLTSAGQSH---- 1052
Db 829 DLDTGTCSLHTASTLDDDDGNYTMAANPQGRISCTGRLMQAVNQGRSPSPSGHPVHR 888
QY 1053 --RGRSRVQERDKPELQERFRPHFLOAPGDMVAHGRCLRCLCKVSGLPPELTWLLNG 1110
Db 889 RPRSRSRDSGDENEPIQERFRPHFLOAPGDLTVQEGKLCRMDCKVSGLPTPDLISWQLDG 948
QY 1111 QVLPDASHKMLVRETGVHSLLDPLTQORDAGTYKCIATNKTNQNSFSLSVVAKEVKK 1170
Db 949 KPRPDSAHKMLVRENGVHSLIIEPTVSRDAGITCTIATNRAGONSFSLVVAKEAHK 1008
QY 1171 APVILEKLQNGVPEGHVPRLECRVIGMPPPVYFKKDNETIPCTREISMHOOTTYGAC 1230
Db 1009 PPVIEKLQNTGVADGYVPRLECRVGLVPPQIEMKKENESLTHSTDRVSMHQNHGYIC 1068
QY 1231 LLIQAPKKSAGWVTTLSAKNEAGIVSCTARLDIYAQWH 1268
Db 1069 LLIQATKEDAGWTVTSKNEAGIVSCTARLDVVISRH 1106
```

```
RESULT 4
Q9Y2J6 PRELIMINARY; PRT; 772 AA.
AC Q9Y2J6;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

```
DE KIAA0992 protein (Fragment).
GN KIAA0992.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023209; BAA76836.1; -.
DR HSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00047; Ig; 3.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00408; IGC2; 3.
DR NCBI_TaxID=9606;
FT NON_TER 1
SQ SEQUENCE 772 AA; 83643 MW; EE5EE2BC79C7492B CRC64;

Query Match 24.2%; Score 1666; DB 4; Length 772;
Best Local Similarity 45.2%; Pred. No. 7.2e-103;
Matches 361; Conservative 113; Mismatches 193; Indels 132; Gaps 20;

QY 621 RQTPDSXQERFNGQATKTPSPVKEPPVLA-----KPKLDSTQLQOLH 667
DB 7 RGKPD-----CPAGPRSLQLP---LAPDAESGSGRRRPEPR-DPLKLOOLQ 53

QY 668 NQVLEHQ-LQNPP---SSPKFFP-----XMTVLNSNAPAVT 704
DB 54 NQIRLEQAGARQPPAPRSAPSPPPPPPPAFPELAACPPASPPEMSALASAPAM- 112

QY 705 TSXKQVAPSSQTSRLAPKYFFSTNTAAT---VAPSSPVFTLSPTQTIQTVSKES 762
DB 113 -----QSSGSFNYPARQKQFIAQNLGPGASGHCTPASSP--SSSLPSPMSTPRQFG 162

QY 763 LLVSHPSVQ-----TKSPGGLSIQNEPLP---GPTPTPTPTFTSPSGNQPPRCVS 813
DB 163 RAVPPPAQPGAPPEAPWSSSPSPPPPPPPVFSPTAAFPVDFPLP-----PPPPPL 217

QY 814 PIP-----VSPTSRI---QNPVAFSLSVLSPFPAIPPTNAXMLPRSAFMPSPQGLAKN 864
DB 218 PSPQASHCSSPAPTFGHGQTAAFLSALLPSQPPPAVNALGLPK---GVTPAGFPKKA 274

QY 865 TKSPQPNNDNIRTKNAVIRDLGKKITFTSDVRPN----- 899
DB 275 SRTARIASDEEIOGTDAVIDQLERLFRKEDLLNNGQPRLTYYERMARRLLGADSATVF 334

QY 900 -----QOEYKISSFEORLWNEIFLERTPVDSEDEIQHDEIPTGKCIAPFEDK 949
DB 335 NIQPEETANQYKVSCEORLISEIYRLERSPVDESDEVGYGVVENGMAFFEM 394

QY 950 RLKHFRTYTESPVFTTKIVGIPVKYWFKDGKQKQISKRNEHCKMRREGDGTCSLHST 1009
DB 395 KLHYKIFEGMPVFTTCRVAGNPKPKIYWFKDGKQISPKSDHYTIQRDLGTCSLHTAS 454

QY 1010 TSDDDGNYTMAANPQGRISCSGHLMVQSLPIRSLRITSAGOSH-----RGRSRVQERDK 1063
DB 455 TLDDDGNYTMAANPQGRISCTGLRMVQAVNQGRSPSPSGHPRRSPRSRSDSDEN 514

QY 1064 EPLQEREFRRHFLQAPGDMVAHEGRCLRLDCKVSGLPPELTLWLLNGPVLDPASHKMLV 1123
DB 515 EPIQEREFRRHFLQAPGDLTVOEGKLCRMCKVSGLPTDLSWLDGKPRVPDSAAHMLV 574

QY 1124 RETGVHSLLDPLTQRTDAGTYKCIATNKTGQNSFSLSVVAKEVKKAPVILEKLQNGV 1183

DB 575 RENGVHSLIIEPVTRDAGIYTCIATNRAGQNSFSLVWAAKEAHPVPFIEKLQNTGV 634
QY 1184 PEGHVPVLECRVIGMPPPVFWKXNDTIPCTRERISHQDQTTGYACLLIQPAKKSAGW 1243
DB 635 ADGYPVRLCRLVGPVPPQIFWKKENESLTHSTDRVSMHQDNHGYICLLIQGATKEDAG 694
QY 1244 YTLSAKNEAGIVCTARLDIYAQHHH--QIPPPMSVPRPSGSRYSGLTSKGLDIFSASSM 1301
DB 695 YTVSAKNEAGIVCTARLDVYTMHQSQSTKPKKVRPSASRYAALSDDGLDIIKAATQP- 753
QY 1302 ESTMVYSCSSRVVSEDEL 1320
DB 754 EANPSHLTLNTALVESEDL 772

RESULT 5
Q9ET54
ID Q9ET54 PRELIMINARY; PRT; 492 AA.
AC Q9ET54;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Actin-associated protein palladin (Fragment).
GN 2410003B16RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RX [1]
RC SEQUENCE FROM N.A.
RA STRAIN=SWISS WEBSTER/NIH;
RX MEDLINE=20391984; PubMed=10931874;
RA Parast M.M., Okey C.A.;
RT "Characterization of palladin, a novel protein localized to stress
RT fibers and cell adhesions.";
RL J. Cell Biol. 150:643-656(2000).
DR EMBL; AF205078; AAC00078.1; -.
DR HSP; P56276; ITLK.
DR MGD; MGI:1919583; 2410003B16RIK.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig-like.
DR InterPro; IPR003606; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00409; IGC; 3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG_like; 1.
DR NCBI_TaxID=10090;
FT NON_TER 1
SQ SEQUENCE 492 AA; 55072 MW; BC59E5B3E3BAEBD5 CRC64;

Query Match 21.2%; Score 1460; DB 11; Length 492;
Best Local Similarity 55.7%; Pred. No. 2.2e-89;
Matches 274; Conservative 78; Mismatches 96; Indels 44; Gaps 5;

QY 872 NDNIRTKNAVIRDLGKKITFTSDVRPN----- 899
DB 2 SDEEIOGTDAVIDQLERLFRKEDLLNNGQPRLTYYERMARRLLGADSANVFNIQPEE 61

QY 900 -----QOEYKISSFEORLWNEIFLERTPVDSEDEIQHDEIPTGKCIAPFEDK 956
DB 62 TAANQEKVSSCEORLISEIYRLERSPVDESDEVQDPVVENATAPFEMKLYKI 121

QY 957 TEGSPVFTTKIVGIPVKYWFKDGKQKQISKRNEHCKMRREGDGTCSLHSTTSSDDGN 1016
DB 122 FEGMPVFTTCRVAGNPKPKIYWFKDGKQISPKSDHYTIQRDLGTCSLHTASTLDDGN 181

QY 1017 YTIMAANPQGRISCSGHLMVQSLPIRSLRITSAGOSH-----RGRSRVQERDK 1070
DB 182 YTIMAANPQGRISCTGLRMVQAVNQGRSPSPSGHPRRSPRSRSDSDENPIQERF 241

QY 1071 FRPHFLQAPGDMVAHEGRCLRLDCKVSGLPPELTLWLLNGPVLDPASHKMLVRETGVHS 1130
```



```
Db 320 TKDAGWTVTSKNEAGIVSCTARLDVYVQHQSQSTKPKKVRPSASRYAALSDOGLDI 379
QY 1295 PSFASMESTWVYSCSSRSVVESEL 1320
Db 380 KAATQP-EANPSHTLTALVVESEL 404

RESULT 8
Q9CWW1 ID Q9CWW1 PRELIMINARY; PRT; 379 AA.
AC Q9CWW1; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2410003B16RIK protein.
GN 2410003B16RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Arai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RL EMBL; AK010350; BAB26871.1; .
DR HSP; P56276; ITLK
DR MGD; MGI:1919583; 2410003B16RIK.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG_like; 1.
DR Immunoglobulin domain.
SQ SEQUENCE 379 AA; 42157 MW; 42A96BE0AAC8F25 CRC64;

Query Match 18.5%; Score 1271.5; DB 11; Length 379;
Best Local Similarity 61.5%; Pred. No. 6.1e-77;
Matches 233; Conservative 61; Mismatches 76; Indels 9; Gaps 4;

QY 950 RLKHFRVTEGSPVFTCKIVGIPVKVYFQKDGKQISKNECHKMRREGDGTCSLHTEST 1009
Db 2 KLHKYKIFEGMPVFTTCRVAGNPKPKIYFQKDGKQISPKSDHYTIQDLDTCSLHATTS 61
QY 1010 TSDDDGNYTMAANPQGRISCSGHLMVQSLPIRSR--LTSAGSHRGRSRVQRD-----K 1063
Db 62 TLDDDGNYTMAANPQGRVSCVTRGLMVQAVNQGRSPSPSGHPHARRPRSRSDSDEN 121
QY 1064 EPQOERFRPHFLQAPGDVMAHEGRCLRDCKYSGLPPPELTWLLNQPVLPDASHKMLV 1123
Db 1123
```

```
Db 122 EPIQERFRPHFLQAPGDLTVQEGKLCRMDCKVSGLPTPDLWSQDGLKPIRPDSAHKMLV 181
QY 1124 RETGVHSLIDPLQTDAGTVKCIATNKTSQSFSLSESVVAKVKKAPVILEKLCNGV 1183
Db 182 RENGVSLLIEPVSVDAGIYTCIATNRAGONSFNLELVAAKRAHRAKAPVEMKZQNGV 241
QY 1184 PEGHPVRLECRVIGMPPPVYFVKKDNETIPCTRERISMHQDQTTGYACLLIQAPKKS DAGW 1243
Db 242 ADGYPVRLECRVSGVPPPIFWKKENESLTHSTERSVMHQDNHGYICLLIQGATKEDAGW 301
QY 1244 YTLSAKNEAGIVSCTARLDIYAQWHH--QTPPPMSVPSRSGRYSLTSKGLDIFSAFSSM 1301
Db 302 YTVSAKNEAGIVSCTARLDVYVQHQPOPTTKPKKVRPSASRYAALS DOGLDIKAATQP- 360
QY 1302 ESTWVYSCSSRSVVESEL 1320
Db 361 EASPSHTLTALNSGLIVESEL 379

RESULT 9
Q98918 ID Q98918 PRELIMINARY; PRT; 4162 AA.
AC Q98918; P87479;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Connectin/titin (Fragment).
GN TITIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST MUSCLE;
RX MEDLINE=96254045; PubMed=8660363;
RA Yaquina H., Ohtsuka H., Kawamura Y., Kume H., Murayama T., Abe H.,
RA Kimura S., Maruyama K.;
RT "A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle
RT connectin/titin reveals its 2 line binding region.";
RL Biochem. Biophys. Res. Commun. 223:160-164(1996).
[2]
RP SEQUENCE OF 1767-1871 FROM N.A.
RX MEDLINE=96365546; PubMed=8769723;
RA Turnacioglu K.T., Mittal B., Sanger J.M., Sanger J.W.;
RT "Partial characterization of zeugmatin indicates that it is part of
RT the 2-band region of titin.";
RL Cell Motil. Cytoskeleton 34:108-121(1996).
DR EMBL; D83390; BAA11908.1; .
DR EMBL; U64829; AAC60019.1; .
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003606; Ig_MHC.
DR Pfam; PF00047; Ig; 33.
DR SMART; SM00408; IGC2; 14.
DR SMART; SM00410; IG_like; 18.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW Immunoglobulin domain.
FT NON_TER 4162 4162
SQ SEQUENCE 4162 AA; 464971 MW; DB0905C907686649 CRC64;

Query Match 12.6%; Score 871.5; DB 13; Length 4162;
Best Local Similarity 23.5%; Pred. No. 8.9e-49;
Matches 327; Conservative 193; Mismatches 507; Indels 367; Gaps 43;

QY 209 SSVPIPIPADTRDNEVNH-----LEQQEAKRREAEQAASEAGG-----DTPFGSSPS 257
Db 30 SGFPVPSVSWRDQVLSAATLPGVQISFSDGRAKLIPSVTEANSGRYTIQATNGSSQA 89
QY 258 SLYEELP---GPPRPTQKLRSREVPEGTFRVQDCTVGIWIPPPQVRYWYCEGLENSPD 314
Db 90 TSTAELLVTAGTAPPNFSQRLQSMRTARQGSQVRLDVRVTGTPVVKFYRGDVEIQSSPD 149
```

QY	1082	MYAHEGLRLCDCKVCSGLPPPELTWLLNGCPVLPDASHKMLVRETGVSHSLLDLPQTQDA	114
Db	1194	VKCSQGTARTFDLKVGRPMPEYTFHNGQQVYNDYTHKIVIKEDGTQSLIIVPAMPEDS	1253
QY	1142	GYKCIATNTGONSFLELSVVAKEVKAPVILEKLQNGVPEGHVPVLECRVIGMPPP	1201
Db	1254	GEWAVIAQNAGKASVTVLSVAKEDLVPRPREVERLRNVSVKESRLHMAVKATGNPNP	1313
QY	1202	VFYWKDKNETI-POTRERISMHDQDTGYACLLLIQPAKKSADGAWYTVLSAKNEAGIVSCSTAR	1260
Db	1314	DIVWLKNSDIIVHPKYPRIRI-EGTKGAAALNTESTARQDAAWYTATATINKAG--RDTRR	1370
QY	1261	LDIYAQWHHQIPPP	1274
Db	1371	CKVNVVEVHAPEP	1384
RESULT 10			
QYIAR9	Q9IAR9	PRELIMINARY;	PRT; 2164 AA.
AC	Q9IAR9;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Structural muscle protein titin (Fragment).		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=HEART;		
RX	MEDLINE=20084964; PubMed=10618168;		
RA	Ayoub J.C., Turnacioglu K.K., Mittal B., Sanger J.M., Sanger J.W.;		
RT	"Targeting of cardiac titin fragments to the Z-bands and dense bodies		
RT	of living muscle and non-muscle cells";		
RL	Cell Motil. Cytoskeleton 45:67-82(2000).		
DR	EMBL; AF159173; AAF35436.1; -;		
DR	InterPro; IPR003598; Ig_C2.		
DR	InterPro; IPR003600; Ig_Like.		
DR	InterPro; IPR003006; Ig_MHC.		
DR	Pfam; PF000047; ig; 10.		
DR	SMART; SM00408; IGC2; 4.		
DR	SMART; SM00410; IG_Like; 6.		
KW	Immunoglobulin domain.		
FT	NON_TER 2164 2164		
SQ	SEQUENCE 2164 AA; 725852C3888755B CRC64;		
Query Match 12.0%; Score 826; DB 13; Length 2164;			
Best Local Similarity 22.0%; Pred. No. 4e-46;			
Matches 351; Conservativity 195; Mismatches 454; Indels 598; Gaps			
QY	209	SSVPIPIADTRDNEVNHA-----LEQQAARRAEQAASEAAG-----DTPPGSSPS	257
Db	30	SGFPVPEVSWYRDQGVLSAATLPGVQISFSDGKALVPSVTEANSGRVTIOATNGSGQA	89
QY	258	SLYTEEPL---GQPRPTQKLRSVPEGTRVOLDICVIGIPPQVPPVWCEKLENSPD	314
Db	90	TSTALLVTAGTAPNFSORLQSMARQGSQVRLVRVTGTPPVVKFYRDGVEIQSSPD	149
QY	315	IHIVQAGNLHSLTAAFEEDTGRYSCFASNIYGTDSAEYIEGVSSDSEGDPNKE-	373
Db	150	FQLQEGDLISLIATAYPEDSGYISVNATNNVGRATSTAELLIQ-----EEEA	PAKKT 205
QY	374	-----EMNRICKPNEVSSP-----PPTS	391
Db	206	KTIIVSTAQISQTRQARIKKIETHSDARSLTSVEMIEGAAQOLPHKAPRMPRP	PTSK 265
QY	392	AVIPAVPQAOHLVAQPRVATIQOCQSPTNYLQGLGKPIIAAPVFTKMLQNL	ASEGL 451
Db	266	SPTPEVITAKAAMARQSPSPVROSPPSVRVRAPTPSPVRSV-----SPAGRI	314

Qy	452	VVFCRVKYGASP--KVEMYREGTLLIEDSPDRILLQKKPRSMABPEEICTILVIAEVFAED	509
Dy	315	STSPIRVPKSPSIPRKAAQVVTPGA--EVLPPWR--OEGYSATAEAQ-----MKETRVST	364
Qy	510	SGCFTCTA-----SNKYG-----TVSSAQHLHVRGNEDLSNGSLSHSANSTTLAAI----	556
Dy	365	SGTEIRTEERMEGRYGLQEQVTISGAAGEVAAGAKERVRRDSEKTAAVATVVAADQAMV	424
Qy	557	-EPOPSPHPHEPVSVEQPKPLEGVLVNHNPE-----RSSRIGRLRVHFNLPEDDKGSE	610
Dy	425	REPAPS-----VVEQ--AAKRTAMTAHVQOVQOEQRKEATV-----VWTSDAKTK	468
Qy	611	----ASSEAGVTTTROTRPDSXQERFNGQATKTPESPFPV-----KEPPPYLAKPLKDST	661
Dy	469	QTVISRTREGIVTSQEORHIS-REKVKRPEKPTPVPTVIATDRAKEQERISTAREEISA	527
Qy	662	QLQOLHNOVLLEHQLONPDPSPSKPEFPFXMTVLNSNAPPA-----VT	704
Dy	528	RHEQVH--VSHEQIRREDVKPSYPK-----VVIITDKPAPVLISKSEGIATKKEHVS	579
Qy	705	TSXKOVKAPSSQTFSLARKPYFFPSNT-----TAATV	737
Dy	580	ISEHKIKKEAKKTVV--PRIAPVTVSTRREEITAKOEQMHLAYDQIEAGKRAEAVTV	637
Qy	738	--APSSSPVFLLSTPQTQIRTVSKESL-----	763
Dy	638	VAAVDQARVSPWETEQVDETYVKKKTYEYGYKEHAVKDHEQAEEHHVATKEVKTVVPPP	697
Qy	764	-----GGLSTONE-----LVSHPSVOTKSP-----	775
Dy	698	EKHIPAAEKKEHVHSTEIKRETEAKTEKIHTHEHRPRTASPHTVSKIAVPKPDPHTYEY	757
Qy	776	-----GGLSTONE-----PLPGPTEP-----TPPPTTSIPS-----	803
Dy	758	SIAGSAMASLEKELSATSQBKITRPVKQPQKLPHQVKIKPESAPPFPFTEAAETVKAH	817
Qy	804	-----GNOFQRCVSPVPVSTSRIQNPVAFLSVLPSPAIPPTN	844
Dy	818	DDVETKEVDVSIKGEAVREDHLLLRKESEAKVTETARPVPA-----EIPVTPPTL	869
Qy	845	AMXLPRSA-----PSMPSQ-----GLAK-----	862
Dy	870	VMGLKNKTVTGESVTLCHISHGPQPTVWYREDYKIESMDQIFTKAGLARLVIREA	929
Qy	863	-----KNTKSPQPVNDQ-NIRETKNAVIRD	886
Dy	930	FAEDSGRFTCATNKAGSVSTSCHLLHVKSVESETRETRESKVTEKSYVETKDVWMED	989
Qy	887	-----LCKK-----	890
Dy	990	VSAAAEEVSGEPPVPFIRKPVVHKLEGGSIIFECQVGGNPKPHVLWKGGVPLTTGYR	1049
Qy	891	-----ITFSQ-----	895
Dy	1050	YKVSYYKRETGECKLEISMTFADDAGGETYIVIRNKFGESATVSLLLEADYEAIKSOQM	1109
Qy	896	-----VRP-----NQOEYKI	905
Dy	1110	MYQTQVTVVQEPKVAEAPPISYGDFOKEYEKEQALLIRKMAKDTVMVVRTFVEDESFHI	1169
Qy	906	SSFQEURLMNEIERLBERTPVD--ESDDIQHDEIPTCGKIAPTDFKRKLKHFRVTEGSPV	962
Dy	1170	SSEFERLIKEIELRIKTIWLDELLEDGEMMIDISEEACGAGFDLRLKNYRTFECTGV	1229
Qy	963	TFTCKIVGIPVPKVMFKOGKOISKRNEHCMMRREGDGTCSLHIESTTSDDDGNYTIMAA	1022
Dy	1230	TFCHKKTGYPLPKIAWYKDGKRI-RHGERYHMEVLQDGSASLRPLVPLPEDEGIYTVFAS	1288
Qy	1023	NPOGRISCSCHLMVQSL-----PIRSRLTSAGQSRRORSR	1057
Dy	1289	NMGNAJCSAKLVKEVPVAPTATPGYNPGEVMRRYRSISRPSRSPARSPCSPAR-R	1347
Qy	1058	VQBRDXEPLQERFRPFHFLQAPCDGMVAHEGRLCRDCKVSGLPPPELTWLLINGOPVLUPDA	1117


```
Db 863 TTVEREFVLHGREAKVTETARPAPVEIPV-----TPPTLVSLGNVTVIEGESV 914
QY SLARPKFFFPSTNTJ----- 733
Db 915 TLECHISGYSPTVWYREDYQIESIDFQITQSGTARLMIREAFEDSGRFTCSAVNE 974
QY AATVAPSS-----SPVFTLSSTPQTQRTVSKESLVSHPVSQTKSPGGLSIQNEPLPP 787
Db 975 AGTVSTSCYLAVQVSEFEKETTAVTEKFTTEKRFVESRDVWMTDT---SLTEE--QA 1028
QY GPTPEPTPPF-----TFSIPSGNQFQRCV---SPIPVSTSR----- 822
Db 1029 GPGEPAAPYFITKPVQKLVGGSGWFGCVQGNPKPHVYWKXSGVPLTTGYRYKVSYNK 1088
QY -----IONVAFLSVLPALPAPPTNAMKLPSPASMPSSQ 859
Db 1089 QTCEKLVISMTFADDAGEYTIIVRNKHGETSA---SASLEEADYELLMKSOEMLYQ- 1144
QY LAKNNTKSPQVNDNIRET-----KNAVIR-DLGKKTTFSDVRPNQOEY 903
Db 1145 ---TQVTAFAVQEPKVGETAPGFGVSEYEKEQALIRKKMAKDTVVVVRTVYVEDQEF 1199
QY KISFEORLWNETFERLERIP-----VDESDEIQHDEIPTGKCTAIPDKRLKHFRVTEG 959
Db 1200 HISSFEERLIKEIYRIKTTLEELLEDEGEKMAVDISESEAVESFGDLRIKNYRILEG 1259
QY SPVTETCKIVGIPVKYVWEKDGKQISKNEHCKMRREGDGTCSLHIESITSDDGNYTI 1019
Db 1260 MGVTFHCKMSGYPLPKTAWYKDGKRI--KHGERYQMDFLQDGRASLRIPVVLPEDEGIYA 1318
QY MAANPOGRISCSGHLVQ-----SL----- 1042
Db 1319 FASNIKGAICSGKLYVEAPALGPATYIPTLEPVSRISLSPRSVSRSPIRMSP 1378
QY SRLTSACQSHGRS---RVQERDKEPQERFRPHFTQAPQMDVMAHGRCLRCDCKVSL 1099
Db 1379 ARMSPARMSPARSGRLEETDESQ--ERLYKPFVFLKPVSEKLEGQATARFDLVVGR 1437
QY PPPELTWLLAGQVLPDASHKMLVRETVHSLDPLTORDAGTYKCIATNKTQNSFSL 1159
Db 1438 PMPETFWFHGQQIVNDYTHKVIKEDGTQSLIIVPATPSDGEWTVVAQNRAGRSSISV 1497
QY ELSVAKEVKKAPVILEKLCNGVPEGHVRLCECRVICMPPPVFYMKDKNETI--PCTRER 1218
Db 1498 ILTVEAVEHQVKPFVEKLNKVNKEGSQLMKVRATGNPNPDIVLKNSDIIVPHKYPK 1557
QY ISMHQDTTGACLLIQAPKSDAGWYTLAKNEAGIVSCTARLDIYAQWHQIOPPPMSVR 1278
Db 1558 IRI--EGTKGEAALKIDSTVSQDSAWYTATATINKAGRDTTRCKNVNVEFAEPEPERKLII 1616
QY PGS 1282
Db PRGT 1620

RESULT 13
O97791
ID O97791 PRELIMINARY; PRT; 2000 AA.
AC O97791;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Titin (Fragment).
GN TTIN.
OC Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SOLEUS SKELETAL MUSCLE;
RX MEDLINE=99034591; PubMed=9817758;
RA Gregorio C.C., Trombitas T., Kolmerer B., Stier G., Granzier H.,
```

```
RA Kunke K., Suzuki K., Obermayr F., Herrmann B., Sorimachi H.,
RA Labelit S.;
RT "The N terminal of titin spans the Z-Disc. Its interaction with a
RT novel 19 kDa ligand (T-cap) is required for sarcomeric integrity.";
RL J. Cell Biol. 143:1013-1027(1998).
DR EMBL; Y18102; CAA77028.1; -.
DR HSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 9.
DR SMART; SM00408; IGC2; 4.
DR SMART; SM00410; IG_Like; 4.
KW Immunoglobulin domain.
FT NON_TER 2000 2000
SQ SEQUENCE 2000 AA; 222702 MW; D157B54C33D5BB68 CRC64;

Query Match 11.5%; Score 793; DB 6; Length 2000;
Best Local Similarity 21.0%; Pred. No. 5,7e-44;
Matches 339; Conservative 193; Mismatches 497; Indels 584; Gaps 46;

QY 198 SFDLSERRERSSVPIPIPADTRDNEVNHAEQQEAKRREAEQAASEAAGDTPGSSPS 257
Db 57 SFS-----GRAKLTPAVTKANSRYSL-----RATNGSGQATSAE 94
QY 258 SLVYEEPLGPPREFTKLRSREVPETGRVQLDCIVVGIPPPQVRYWYCEGKELENSPDIHI 317
Db 95 LLVTAET--APPNVQRLQSWTVRQGSQVRLQVRVTGPTPVVKFYKDGAEIQSLDFQI 152
QY 318 VQAGNLSLTIAEAFEDTGRYSFASNIYGTDSIAEYIEG-----VSSDSEGD 369
Db 153 SQEGDLSLLIAEAYPEDSGTYSVNATNSVGRATSTAEALLVQGEVVPKAKTKTIVSTAQ 212
QY 370 PKEMNRIOK-----PNEVSPPTTSIVIPAVP 399
Db 213 ISETROTRIEKKIBAHFDARSIAIVMVVDGATQQLPHKTPPRIPKPKRSRSTPPSIA 272
QY 400 QAQLHVAQPRVATIQCCQSPNTNYLQGLDGPI-----IAAPVFTKMLQ 443
Db 273 AKAQLARQSPSPRHSPSPVRHVRAPTPSPRSVSPAGRISTSPIRSKSPLLVRKTQT 332
QY 444 LSASEGOLVFECEKRVKSGAPSKVEYREG-----TLIEDSPDFRI----- 483
Db 333 TTMATPEV-----PPP---WKQEGYVASSTEAEEMRETTWTSSEIRREERWEGRY 380
QY 484 -LQKKPRSMABPEICILVIAEVAEDSGCCTCTA-----SNKYGVSSI--AQLHVRG 534
Db 381 GLQEQVITGAAGAAAAVSTSAAFAGLGSWFATGAKEVKQADKSAAVATVVAADVMAV 440
QY 535 NEDLSNGLSHSANTTNLAAIEPOPSPPHSEPPSV-----EQPK----- 575
Db 441 REPVISAVEQTAQRTTTTAVHIQPAEQVRAEKIAVSVVVAADKAKEQELKSRPREI 500
QY 576 --PKLEGVLVNHNEPR-----PVKEP-----PPVLAKPK---LDSTQL 663
Db 501 ITTKQEQMHITHEQLRKETEKAFFVKPVVISAAKAKEQETRITGETTTKQEQKITQETMM 560
QY 590 -----SSRIGL---RVHFNLPEDDKGSEASSEA-----GVV 618
Db 561 KETRKTVPVKIVATPKVKEQDLVSRREGISTKREQVQITQERMKRAEAKTALSTIATA 620
QY 619 TTRTRPDSXQERFNGQATK-----T 639
Db 621 TAKATEQETVLTREGMATROEQIQVTHGKVGKGAEAFAVATVVAADVQARVREPREG 680
QY 640 PEPSF-----PVKEP-----PPVLAKPK---LDSTQL 663
Db 681 PEDSYAQOTTLEYGYKEHISATKVAEHPKRPASEPHVVPKAVKPVVQIAPSETHIKTTEQ 740
QY 664 QQLH-----NOVLLEQ---HQLQNPPSPSKPEFPFXMTVLN-----SNAPPV 703
Db 741 VGMHISQIKKTTDVTSERLVHDKRPRTAGSPH---FTVSKISVPKTEHGEYASAGSAI 797
```


Db 633 ITTKREOVQITQEMRKEAEKTAISTALSTAVATAKAEQETILTRRETMATRQEIQVTHGK 692
 QY 517 -----ASNKYGVTSIAQLHVRGNEEDLSNNGSLHSANST-----TNLAAIEPQSP 562
 Db 693 VDVGKAAEAVATVAAQDQARVREPREGHLEESYAQOTTLEYGYKERISAAKVAEPQOR 752
 QY 563 PHSEPPSVEQPPKPL-----EGVLVNHNEPRSS 591
 Db 753 PASEPHVVPKAVPRVIAQSETHIKTTDQGMHISSQIKKTTDLTERLVHVDKRRPTA 812
 QY 592 SRIGLRVHF-----NLPEDDKGSEASEAGVVTTRQ-----TRPD 626
 Db 813 S-----PHFTVTSISVPKTEGHGEASIAIATLQELSATSAAQITKSVKAPTVPKS 867
 QY 627 SXQERNGQATKTPESFPKPEPPVLAKP-----KLDSTQLQOLHNOVL--LE 673
 Db 868 ETRVR-----AEPPLPQFPADPTDYKSEAGVEVKEVGVSTGTTVREEREFEVLHGRE 923
 QY 674 QHQLQNPSPSPKREFFFXMTVLNSNAPPAVTTTSXQVKAPSSQTSFLARPKYFPFSTNTT 733
 Db 924 AKVTEARVAPVEIPV-----PPPTLVSLGKNVTVIEGESVILECHISGYPSTVT 975
 QY 734 -----AATVAPSS-----S 742
 Db 976 WYREDYQIESSIDFQITFQSGIARLMIREAFEDSGRFTCSAVNEAGTVSTSCYLAVOVS 1035
 QY 743 PVFTLSTPOTIORTVSKESLLVSHPSVQTKSPGGLSIQNEIPLPGPTETPTPPPF----- 797
 Db 1036 EEFKEKETTAVTEKFTTEERFVESRDVVMYDT-----SLTEE--QAGPGEPAAFPYFITKP 1089
 QY 798 -----TFSPISGNQOPRCV---SPIPVSPTSR----- 822
 Db 1090 VQKLVEGVSVFCQVGGNPKPHVYKKKSGVPLTTGYRYKVSYNKQTCGEKLVISMTFAD 1149
 QY 823 -----IQNVAFITSSVLPSPALPPTNAMXLPSPAPSMPSOGLAKNKTSPQPVNDD 874
 Db 1150 DAGEYTTIVRNKHGETSA---SASLLEEADYELLKMSQOEMLYQ-----TQVTAFAVQEP 1200
 QY 875 NIRET-----KNNAVIR-DLGCKKITFSFDRVNRQOEYKTISSFEQRLMNEIEF 918
 Db 1201 KVGETAPGVVSEYEYKEQALIRKKMAKQTVVNTYVVEDQEFHISSEERLKEIEY 1260
 QY 919 RLERTP-----VDESDDIEQDEIPTGKCIAPIFDKRLKHFVRTEGSPVTETKIVGIPVP 974
 Db 1261 RIUKTILEELLEEDGEKMAVDISESEAVESGFDRLKNVRILEGMGVTFHCKMSGYPLP 1320
 QY 975 KVIWFDGKOISKRNEHCKMRREGDGTCSLHIESTTSDDDGNTYTIMAANPQGRISCSGHL 1034
 Db 1321 KIAWYKDGKRI-KHGERYQMDFLQDGRASLRIPVPLPEDEGIYTAFAFNKNAICSGKL 1379
 QY 1035 MVQ-----SL-----PIR-----SRLTSAGQSHRGSR- 1056
 Db 1380 YVPAAPLAPGAPTYIPTLEPVRSIRSLSPRSVRSPIRMSPARMSPARMSPARMS 1439
 QY 1057 --RVQERDKRPLQERFRPHFLQAPGDMVAHEGRCLRDLCKVSGLPPELTLLNGOPVL 1114
 Db 1440 GRLEETDESQLE-RLYKPVFLKPVSKFCLGQTARFDLKVVGPRNPPEFWFHGCGQIV 1498
 QY 1115 PDASHKMLVRETGVHSLLDPLTQDAGTYKCIATNKTGONSFSLELSVVAKEVKKAPVI 1174
 Db 1499 NDVTHKVVIKEDGTQSLIIVPATPSDGEWTVVAQNRAGRSSISVILTVEAVEHVKPMF 1558
 QY 1175 LEKLQNGVPEGHVRLRECVIGMPVPVFWKKDNETI-PCTRERISMHDQTTGYACLLI 1233
 Db 1559 VEKLKNVIRKEGQLEKMKVATGNPNPDIIYVLKNSDIIIVPHKYPKIRI-EGTGERALKI 1617
 QY 1234 QPAKKSAGWYTTLSAKNEAGIVSCTARLDIYAOWHHOIPPPMVSVRPSGS 1282
 Db 1618 DSTVSDSAWYATATANKAGRDTRCKVNVNEFAEPEPERKLIIPRGT 1666

RESULT 15

Q8WZ42
 ID Q8WZ42 PRELIMINARY; PRT; 34350 AA.
 AC Q8WZ42;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE titin.
 GN TTN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20309627; PubMed=10850961;
 RA Freiburg A., Frombitas K., Hell W., Cazorla O., Fougereousse F.,
 RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
 RA Granzier H., Labeit S.;
 RT "Series of exon-skipping events in the elastic spring region of titin
 as the structural basis for myofibrillar elastic diversity.";
 RL Circ. Res. 86:1114-1121(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21573839; PubMed=11717165;
 RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
 RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,
 RA Labeit S.;
 RT "The complete gene sequence of titin, expression of an unusual ~700
 kDa titin isoform and its interaction with obscurin identify a novel
 Z-line to I-band linking system.";
 RL Circ. Res. 89:1065-1072(2001).
 DR EMBL; AJ277892; CAD12456.1; -;
 DR InterPro: IPR000282; Cytok_receptor_2.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000577; FGGY_kin.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001092; HLH_basic.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_v.
 DR InterPro: IPR002016; Peroxidase.
 DR InterPro: IPR004168; PPAK_motif.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam; PF00041; fn3; 132.
 DR Pfam; PF00047; Ig; 146.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02818; PPAK; 53.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00060; FN3; 133.
 DR SMART; SM00409; IG; 167.
 DR SMART; SM00408; Igc2; 148.
 DR SMART; SM00406; IGV; 23.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00933; EGGY_KINASES_1; UNKNOWN_1.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
 SQ SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;

Query Match 11.3%; Score 780.5; DB 4; Length 34350;
 Best Local Similarity 20.0%; Pred. No. 2e-41;
 Matches 333; Conservative 205; Mismatches 488; Indels 643; Gaps 42;

QY 198 SFDLSRRSSVPIPIPADTRNEVNHAEQOEAKRREAEQAAGDTPPGSSPS 257
 Db 57 SFSD-----GRAKLTIPAVTKANSGRYSL-----RATNGSGQATSTAE 94
 QY 258 SLIYEELGQPPRPTQKILRSREVPEGTREVOLDICIVGIPPPQVHWYCEGKELENSPDIIH 317

Db 95 LLVKAET--APPNVQRLQSMVTRQSGOVRLQVRVTGIPTPVVKFYRDGAIEIOSSDFQI 152
Qy 318 VQAGNLHSLTIAEAFEDTGRYSCFASNIYGTDSIAEYIEG-----VSSSDSEGD 369
Db 153 SOEGDLSLLIAEAYPEDSGTYSYNATNSVGRATSAELLVQOGEVEEPAKTKIIVSTAQ 212
Qy 370 PNKEEMRIQK-----PNEVSSPPTTSVAIPPAVP 399
Db 213 ISESRQTRIEKKIEAHFDANSIATVEMVIDGAAGQQLPKHTPHRIPPKPKSRSPTPPSIA 272
Qy 400 QAHLVAPRVATTOQCOSPNTYLOGLDGKPI-----IAAPVETKMLQN 443
Db 273 AKAQARQOQSPSRHSPSPVHRVRAPTPSVRSVSPAARISTSPIRSRSPLLMRRTQA 332
Qy 444 -----LSASEGOL-----VVFECEVRKGA 462
Db 333 STVATGPEVPPKQEGVAVSSSEAEEMRETTLTSTQIRTEEREGYGVQEQVTISGAA 392
Qy 463 S----- 463
Db 393 GAAASVSASAYAAEAVATGAKEVKQADKSAAVATVVAADMARVREPVISAVEQTAQR 452
Qy 464 ----- 463
Db 453 TTTTAVHIQPAQOVRKEAEKTAFTKVVAADKAKEQELKSRTKEVITTKOEQMHVTHEQ 512
Qy 464 -----PKV-----EWYRE----- 471
Db 513 IRKETETFPVKVISAAKAKEQETRISEETKKOKQVTOGAIRQETETIAASWVVVATA 572
Qy 472 -GTLIEDSP----- 479
Db 573 KSTKLETVPGAQEBETTQOQOMHLSYEKIMKETRTVVPKVIVATPKVKEODLYSRGREG 632
Qy 480 -----DFRILQKPRMAEPEETCTLVIAEVAEDSGCFTCT----- 516
Db 633 ITTKREQVITQEKMRKEAEKALTALSTAVATAKAKEQETILRTRETMATROEQIQVTHGK 692
Qy 517 -----ASNKYGTVSSIAOLHVRGNEDLSNNGSLHSANST-----TNLAATEPOPS 562
Db 693 VDCKKKAENATVVAADVQARVREPREGHLEESYAQOTILEYGYKERISAAKVAEPPQR 752
Qy 563 PHSEPPSVEQPPKPL-----EGLVNHNEPRSS 591
Db 753 PASEPHVVPRAVKPRVIAQAPSETHIKTTDQGMHISQIKKTTDLTTERLVHVDKRPRTA 812
Qy 592 SRIGLRVHF-----NLPEDDKGSEASSEAGVVTTRQ-----TRPD 626
Db 813 S-----PHFTVSKISVPKTEHGYEASIAIATLOKELSATSQAQITKSVKAPTVPKS 867
Qy 627 SXQERENGQATKTPERSFPYKPPVLAKEP-----KLDSTOLQQLHNOVL--LE 673
Db 868 ETRVR-----AEPTLPQOPFPADTPTDTKSEAGVEKKEVGSITGTIVREFEVLHGRE 923
Qy 674 QHQLQNPSPSPKPFPPXMTVLNSNAPPVAVTTSXQVKAPSSQTFSLARPKYFPFSTNTT 733
Db 924 AKVTETARVPAPVEIPV-----TPPTLVSGLKNVTVIBGESVTLCHISGYPSPTVT 975
Qy 734 -----AATVAPSS-----S 742
Db 976 WYREDYQIESIDFQITFQSGIARLMIREAFEDSGRFTCSAVNEAGTVSTSCYLAVQVS 1035
Qy 743 PVFTLSSTPOTIORTVSKESLLVSHSPSVQKSGGLSTQNEPLPPGTEPTPPPF----- 797
Db 1036 EEFKEKETTAVTEKETTEKREKRESDRVVNDT-----SLTEB--QAGGEPAAPIFTKPV 1089
Qy 798 -----TFSIPSGNQFQPCV-----SPIPVSPTSR----- 822
Db 1090 VOKLVEGSGVVGCGVGNPKPVYKKGVPVLTGTGYRYKVSYNKQGTGECKLVISMFTAD 1149
Qy 823 -----IQNPVAFLSVLPSPAIPTTNAMXLPRAAPSMPQSGLAKKNTKSPQPVND 874

Db 1150 DAGETTVIVRNKHGETSA---SASLLBEADYELLMKSQOEMLYQ-----TQVTAFTVOEP 1200
Qy 875 NIRET-----KNAVIR-DLGKKITFSDVRPNQOQYKISSFQRLMNIETEF 918
Db 1201 KVGETAGFVYSEYEKEQEALIRKKMAKDTVVVRYVEDQEFHISSEERLIKIEY 1260
Qy 919 RLERTP-----VDESDDIEQHDEIPTGKCIAPIFDKRLKHFRVTEGSPVTTCKTIVGIPVP 974
Db 1261 RIITKTTLEELLEEDGEEKMAVDISESEAVESGFDLRINKRYRILEGMGVTEHCKMSGYPLP 1320
Qy 975 KYVWFKGQKQISKNEHCKMRREGDGTCSLHISTTDDGNYTIMAANQGRISCSGHL 1034
Db 1321 KIAWYKDGKRI-KHGERYQMDFLQDGRASURIPVVLPEDEGIYTAFAASNKGNAICSGKL 1379
Qy 1035 MVQ-----SL-----PIR-----SRLTSAGQSHRGSR- 1056
Db 1380 YVEPAAPLGAPTYIPTLEPVSRIRSLSPRSVSRSPIRMSPARMSPARMSPARMS 1439
Qy 1057 --RVQERDKEPLOERFPRPHFLQAPGDMVAHEGRCLRCLCKVSGLPPPELTWLLNGQPV 1114
Db 1440 GRRLEETDESOL-ERLYKPVFVLKPVSKLEGGOTARFDLKVVGRRPMPETFWPHDGOQIV 1498
Qy 1115 PDASHKMLVRETGVHSLLDPLTORDAGTYKCIATNKTGQNSFSLSELSVVAKEVKKAPVI 1174
Db 1499 NDYTHKVVIKEDGTQSLIIVPATPSDSGEWTVVAQNAGRSSISVILTVEAVEHQVKPMF 1558
Qy 1175 LEKLONGCVPEGHVPVRLCECRVIGMPPVPVYWKDNETI-PCTREIRISMHODTTGYACLLI 1233
Db 1559 VEKLNKYNIKEGSOLEMKVRATGNPNPDIVWLKNSDIIVPHKYPKIRI-EGTKGEAALKI 1617
Qy 1234 QPAKKSADAGWYTLTSAKNEAGIVSCTARLDIYAQWHHQIPPPMSPVRPSGS 1282
Db 1618 DSTVSQDSAWYTATAIKAGRDTTRCKVNVVEFEAEPEPERKLIIPRG 1666

Search completed: November 30, 2002, 19:06:18

Job time : 53 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 18:50:40 ; Search time 37 Seconds
(without alignments)
4753.806 Million cell updates/sec

Title: US-09-818-990b-2
Perfect score: 6890
Sequence: 1 MQDDIEASTISQLRESY.....MESTWVSCSSRSVDESDEL 1320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				A_Geneseq_101002:*	
Database :				1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
Database :				2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
Database :				3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
Database :				4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
Database :				5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
Database :				6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
Database :				7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
Database :				8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
Database :				9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
Database :				10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
Database :				11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
Database :				12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
Database :				13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
Database :				14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
Database :				15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
Database :				16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
Database :				17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
Database :				18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
Database :				19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
Database :				20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
Database :				21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
Database :				22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
Database :				23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	2560	37.2	507	22	Human protein sequ
2	2174	31.6	408	22	Novel human diagno
3	1477	21.4	544	21	Human colon cancer
4	1421.5	20.6	433	21	Human ORFX ORF2526
5	1394	20.2	264	22	Human cDNA SEQ ID
6	1380	20.0	261	22	Human immunoglobul
7	797.5	11.6	26926	22	Human titin (conne
8	729	10.6	498	21	Amino acid sequenc
9	724.5	10.5	159	22	Human cDNA SEQ ID
10	724.5	10.5	159	22	Human immunoglobul

11	656	9.5	130	21	AA040875	Human ORFX ORF639
12	556	8.1	1953	23	AA084351	Protein MYLK diffe
13	545.5	7.9	4796	22	AB058665	Drosophila melanog
14	537	7.8	5635	23	AB060991	Novel human protei
15	531.5	7.7	4412	21	AA053666	Sequence gi/101742
16	475	6.9	3117	21	AA053667	Sequence gi/332818
17	472	6.9	6815	22	AB066811	Drosophila melanog
18	462.5	6.7	1240	22	AB039333	Novel human diagno
19	446.5	6.5	201	23	AA074341	Human cytoskeleton
20	426	6.2	7107	22	AB058144	Drosophila melanog
21	416.5	6.0	2597	21	AA053664	Mechanical stress
22	416.5	6.0	2597	23	AA047933	Rat OCP. Rattus r
23	410	6.0	2587	23	AA047935	Human OCP. Homo s
24	408.5	5.9	2387	21	AA053665	Mechanical stress
25	392	5.7	4393	22	AB031889	Amino acid sequenc
26	392	5.7	4436	22	AB023265	Novel human diagno
27	390	5.7	1853	21	AA053668	Protein 608 sequen
28	388	5.6	1114	22	AB066628	Drosophila melanog
29	387	5.6	2993	22	AB023884	Novel human diagno
30	384	5.6	5266	22	AB08561	Novel human diagno
31	383	5.6	3263	22	AB067210	Drosophila melanog
32	381	5.5	2286	22	AB065635	Novel protein kina
33	381	5.5	2380	23	AA019160	Human kinase poly
34	377	5.3	913	22	AB059114	Drosophila melanog
35	367.5	5.3	162	23	AB07380	Human ORFX protein
36	353.5	5.1	2542	22	AB071137	Drosophila melanog
37	353	5.1	2016	22	AB066424	Drosophila melanog
38	339	4.9	1336	23	AB047771	Human thrombospond
39	334.5	4.9	2596	22	AB030569	A splice variant o
40	332	4.8	681	22	AB057982	Drosophila melanog
41	321.5	4.7	1297	20	AA013565	C. elegans Robo po
42	321.5	4.7	1297	20	AA08403	C. elegans Robo pr
43	318.5	4.6	1395	20	AA013563	Drosophila Robo 1
44	318.5	4.6	1395	20	AA08401	Drosophila sp. ROB
45	317	4.6	1465	23	AA016625	Human M-protein en

ALIGNMENTS

RESULT 1	
AA092617	
ID	AA092617 standard; Protein; 507 AA.
XX	
AC	AA092617;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:10896.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	(HELI-) HELIX RES INST.
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 10896; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 507 AA;

Query Match 37.2%; Score 2560; DB 22; Length 507;
Best Local Similarity 100.0%; Pred. No. 3.8e-145;
Matches 494; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQDDSI EAS TIS IOLLRES YLAETFRHGN NERSRAE PSSN PCHFGSPSGAAEGGGQDDL 60
Db 1 MQDDSI EAS TIS IOLLRES YLAETFRHGN NERSRAE PSSN PCHFGSPSGAAEGGGQDDL 60
QY 61 PDL S A F L S Q E L D E S V N L A R L A I N Y D L E K A D E T Q A R K L S P D Q M K H S P N L S F P E P N C Q D 120
Db 61 PDL S A F L S Q E L D E S V N L A R L A I N Y D L E K A D E T Q A R K L S P D Q M K H S P N L S F P E P N C Q D 120
QY 121 N P R S P T S S K E S P Q E A K R P Q C S E T Q S K K V F L N K A A D F I E E L S S L F K S H S K R I R P R A C K N 180
Db 121 N P R S P T S S K E S P Q E A K R P Q C S E T Q S K K V F L N K A A D F I E E L S S L F K S H S K R I R P R A C K N 180
QY 181 H K S K L E S Q N K V M Q D N S S F S D L S R R E R S S V P I P A D T R D N E V N H A L E Q Q E A K R R A E Q 240
Db 181 H K S K L E S Q N K V M Q D N S S F S D L S R R E R S S V P I P A D T R D N E V N H A L E Q Q E A K R R A E Q 240
QY 241 A A S E A A G D T T P G S P S S L Y E E P L G O P P R P T Q K L R S R E V P E G T R V O L D C I V G I P P Q V 300
Db 241 A A S E A A G D T T P G S P S S L Y E E P L G O P P R P T Q K L R S R E V P E G T R V O L D C I V G I P P Q V 300
QY 301 R W Y C E G L E N S P D I H V Q A G N L H S L T A I A F A E D T G R Y S C F A S N I Y G T D S T S A E I Y I E G 360
Db 301 R W Y C E G L E N S P D I H V Q A G N L H S L T A I A F A E D T G R Y S C F A S N I Y G T D S T S A E I Y I E G 360
QY 361 V S S D S E G D P N K E M N R I Q P N E V S S P T T S A V I P P A V P Q A H L V A O P R V A T I Q C C S P T 420
Db 361 V S S D S E G D P N K E M N R I Q P N E V S S P T T S A V I P P A V P Q A H L V A O P R V A T I Q C C S P T 420
QY 421 N Y L Q L D G K P I I A P V T K M L Q N L S A S E G O L V F E C R V K G A P S P K V E W Y R E G T L I E D S P D 480
Db 421 N Y L Q L D G K P I I A P V T K M L Q N L S A S E G O L V F E C R V K G A P S P K V E W Y R E G T L I E D S P D 480
QY 481 F R I L Q K P R S M A E P 494
Db 481 F R I L Q K P R S M A E P 494

RESULT 2
ABG10611
ID ABG10611 standard; Protein; 408 AA.
XX
AC ABG10611;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10602.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS74798.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 40970; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 408 AA;

Query Match 31.6%; Score 2174; DB 22; Length 408;
Best Local Similarity 99.5%; Pred. No. 3.7e-122;
Matches 406; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 913 M N E I E F R L E R T P V D E S D D E I Q H D E I P T G K C I A P I F D R K L K H F R V T E G S P V T F T C K I V G I P 972
Db 1 M N E I E F R L E R T P V D E S D D E I Q H D E I P T G K C I A P I F D R K L K H F R V T E G S P V T F T C K I V G I P 60
QY 973 V P K Y W F K D G K Q I S K R N E H C K M R R E G D T C S L H I E S T S D D G N Y T I M A A N P Q R I S C S G 1032
Db 61 V P K Y W F K D G K Q I S K R N E H C K M R R E G D T C S L H I E S T S D D G N Y T I M A A N P Q R I S C S G 120

thrombosis; contraceptive.
Homo sapiens.
WO200058473-A2.
05-OCT-2000.
31-MAR-2000; 2000WO-US08621.
31-MAR-1999; 99US-0127607.
02-APR-1999; 99US-0127636.
05-APR-1999; 99US-0127728.
30-MAR-2000; 2000US-0540763.
(CURA-) CURAGEN CORP.
Shinkets RA, Leach M;
WPI; 2000-602362/57.
N-PSDB; AAC76971.
Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
Claim 11; Page 4239-4240; 5507pp; English.
AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnary;
antiparkinsonian; neurotropic; neuroprotective;
osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antithyroid; and antianaemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.
Sequence 433 AA;
Query Match 20.68; Score 1421.5; DB 21; Length 433;
Best Local Similarity 61.08; Pred. No. 4.7e-77;
Matches 261; Conservative 69; Mismatches 89; Indels 9; Gaps 3;
QY 901 QEYKISSFEORLMEIEFLERTPVDSDDIEIODEITGKCIAPIDFKRLKHFRVTEGS 960
Db 7 QEYKVSCEQLRIEIEFLERSPVDSGGDEFTYGVDPVNGMAPFFEMKLKHKIIEGM 66
QY 961 PVTFTCKIVGIPVKYVFKDGKQISKRNECHKMRREGDGTCSLHIESTTSDDDGNYTIM 1020
Db 67 PVTFTCRVAGNPKKIYFKDGKQISPKSDHYTIQRDLDTGTCSLHTTASTLDDGNYTIM 126
QY 1021 AANPQGRISCSGLHWOSLPIRSLTISAGQSH-----RGRSRVQERDKPELQERFFRPH 1074
Db 127 AANPQGRISCTGRMLVQAVNGRSPSGPHVVRPRRSRSDSGDENEPQERFFRPH 186
QY 1075 FLOAPGDMWAEGRCLRDCKVSLGPPPELTWLLNGOPVLPLDASHKMLVRETGVHSLIID 1134
Db 187 FLOAPGDLTVOEGKLCRMDCVKSLGPTDLSNQDGLGKVPVPSDAHKMLVRENGVHSLIIE 246
QY 1135 PLTORDAGTYKCIATNKTGONSFSLELSVVAKEYKKAPVILEKLQNCVGPBGHPVLECR 1194

Db 247 PVTSRDAGIVTCIATNRAGQNSFSELVVAAKEAHKPPVFTEKLTQNTGVADGVPVLECR 306
QY 1195 VIGMPPPVFWKKDNETIPTCTRERISMHDYTGACALLIQPAKSDAGWYTLSSAKNEAGI 1254
Db 307 VLGVPFPQIFWKKENESLTHSTDRVSMHQDNHGYICLLIQGATKEDAGWYTVSAKNEAGI 366
QY 1255 VSCARLDIYAOWHH--QIPPPMSVRPSPGSGRYSLTSGKLDIFSAFSSMESTMYVSCSSR 1312
Db 367 VSCARLDVYTOHQOQSOSTKPKKVRFSASRYAALSQGLDIIKAQFP-EANPSHLTLNT 425
QY 1313 SVSEDEL 1320
Db 426 ALVESEDL 433
RESULT 5
ABB10330
ID ABB10330 standard; Protein; 264 AA.
AC ABB10330;
XX
XX 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 638.
XX Human; gene therapy; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation.
XX Homo sapiens.
XX WO200154474-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01349.
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 24-FEB-2000; 2000US-184664P.
XX 02-MAR-2000; 2000US-186350P.
XX 16-MAR-2000; 2000US-189874P.
XX 17-MAR-2000; 2000US-190076P.
XX 18-APR-2000; 2000US-198123P.
XX 19-MAY-2000; 2000US-205515P.
XX 07-JUN-2000; 2000US-214886P.
XX 30-JUN-2000; 2000US-215135P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216880P.
XX 11-JUL-2000; 2000US-217487P.
XX 14-JUL-2000; 2000US-217496P.
XX 26-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225213P.
XX 14-AUG-2000; 2000US-225214P.
XX 14-AUG-2000; 2000US-225266P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225288P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 18-AUG-2000; 2000US-225759P.
XX 18-AUG-2000; 2000US-226279P.
XX 22-AUG-2000; 2000US-226681P.
XX 22-AUG-2000; 2000US-226686P.
XX 22-AUG-2000; 2000US-227182P.

PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 03-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234937P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.

PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR N-PSDB; ABA06552.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
PS Claim 11; SEQ ID NO: 638; 859pp + Sequence Listing; English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
SQ Sequence 264 AA;

Query Match 20.2%; Score 1394; DB 22; Length 264;
Best Local Similarity 99.6%; Pred. No. 1.1e-75;

Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1059 QERDKPELQERFRPHFLQAPGDMVAHEGRLCDCKVSLPPPELFWLLNGQVLPDAS 1118
Db :|||||

3 RERDKPELQERFRPHFLQAPGDMVAHEGRLCDCKVSLPPPELFWLLNGQVLPDAS 62
Qy 1119 HKMLVRETGVHSLIDPLTORDAGTYKCIATNKTKGNSFSLSVWAKEVKAPVILEKL 1178
Db :|||||

63 HKMLVRETGVHSLIDPLTORDAGTYKCIATNKTKGNSFSLSVWAKEVKAPVILEKL 122
Qy 1179 QNCGVPEGHVPRLCECRVIGMPVPVYWKKNDETICTPCTRERISMHDQTTGYACLLIQPAKK 1238
Db :|||||

123 QNCGVPEGHVPRLCECRVIGMPVPVYWKKNDETICTPCTRERISMHDQTTGYACLLIQPAKK 182
Qy 1239 SDAGWYTLAKNEAGIVSCTARLDIYAQWHHQIPPPMSVRPSGSRGSLTSGKGLDIFS 1298
Db :|||||

183 SDAGWYTLAKNEAGIVSCTARLDIYAQWHHQIPPPMSVRPSGSRGSLTSGKGLDIFS 242
Qy 1299 SSMESTMVYSCSSRSRVESDEL 1320
Db :|||||

Db 243 SMESTMVYSCSRVSVESDEL 264
RESULT 6
AAU17989
ID AAU17989 standard; Protein; 261 AA.
XX AC AAU17989;
XX DT 07-NOV-2001 (first entry)
XX DE Human immunoglobulin polypeptide SEQ ID No 134.
XX KW Immunoglobulin; signal transduction pathway protein; cancer;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX OS Homo sapiens.
XX PN WO200155315-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01326.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.


```
PR 14-SEP-2000; 2000US-233400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246529P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.

PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476161/51.
XX N-PSDB; ABA06731.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
XX polypeptide is used in preventing, treating or ameliorating a medical
XX condition
XX
XX Claim 11; SEQ ID NO: 817; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
XX DNAs. These can be used in the treatment of neural, immune system,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal and proliferative disorders and inflammation. The present sequence
XX is a protein of the invention.
XX
XX Sequence 159 AA;
SQ
Query Match 10.5%; Score 724.5; DB 22; Length 159;
Best Local Similarity 89.6%; Pred. No. 6.9e-36;
Matches 138; Conservative 3; Mismatches 10; Indels 3; Gaps 2;
QY 1142 GTYKCIATNKTGQNSFSLSVAKKAPVILEKQNCVPGCHPVRLRCRVIGMPPP 1201
Db 8 GTYKCIATNKTGQNSFSLSVAKKAPVILEKQNCVPGCHPVRLRCRVIGMPPP 67
QY 1202 VFYWKDNETIPTCTRERISMHDYTGACLLIQPAKSDAGWYTLAKNEAGIVSCTARL 1261
Db 68 VFYWKDNETIPTCTRERISMHDYTGACLLIQPAKSDAGWYTLAKNEAGIVSCTARL 127
QY 1262 DIYAQWHHQIIPP-MSVRPSGRYSGLTSKGLDI 1294
Db 128 DIYAQWHHQIIPPCLPAQXQXSRISN--SKGLDI 159
RESULT 10
AAU18053
ID AAU18053 standard; Protein: 159 AA.
XX
XX AAU18053;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immunoglobulin polypeptide SEQ ID No 198.
XX
XX Immunoglobulin; signal transduction pathway protein; cancer;
XX antisense therapy; gene therapy; neurological disorder; renal disorder;
XX cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
XX reproductive disorder; immune system disorder; proliferative disorder;
XX muscular disorder.
XX
XX Homo sapiens.
XX
XX WO200155315-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01326.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
```


XX Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
CC invention. The polypeptides and their associated polynucleotides can be
CC used to diagnose a pathological condition or a susceptibility to a
CC pathological condition in a subject by determining the presence or
CC absence of a mutation in a DNA sequence or determining the presence or
CC amount of expression of the protein. Alternatively the identification of
CC a binding partner to a sequence allows determination of changes in
CC protein activity. The sequences can be used as research tools for
CC receptors or other signal transduction pathway proteins that interact
CC with the polypeptides of the invention and can be used to treat, prevent
CC or diagnose various types of disorders such as neurological disorders,
CC cardiovascular disorders, gastrointestinal disorders, reproductive
CC disorders, immune system disorders, renal disorders, muscular disorders,
CC pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 159 AA;

Query Match 10.5%; Score 724.5; DB 22; Length 159;
Best Local Similarity 89.68; Pred. No. 6.9e-36;
Matches 138; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

QY 1142 GTYKCIATNKTQNSFSLSVAKVKKAPVILEKLCQNGVPEGHVRLRCRVIGMPPP 1201

Db 8 GTYKCIATNKTQNSFSLSVAKVKKAPVILEKLCQNGVPEGHVRLRCRVIGMPPP 67

QY 1202 VFYKKNDNETPCTRERISMQDITGYACLLIQPAKKSADGWTLSAKNEAGIVSCTARL 1261

Db 68 VFYKKNDNETPCTRERISMQDITGYACLLIQPAKKSADGWTLSAKNEAGIXSCTARL 127

QY 1262 DIYAQWHHQIIPP-MSVAPSGSRVSGLSKGLDI 1294

Db 128 DIYAQWHHQIIPXCLPAQXQXSRISN--SKGLDI 159

RESULT 11

AA040875

ID AAB40875 standard; Protein; 130 AA.

XX AAB40875;

AC AAB40875;

DT 08-FEB-2001 (first entry)

XX Human OREF639 polypeptide sequence SEQ ID NO:1278.

XX Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

PR

PR 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

PI WPI; 2000-602362/57.

XX N-PSDB; AAC75084.

PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PS Claim 11; Page 1115-1116; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,

CC which represent the human OREFX open reading frames 1 to 3161. The OREFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an OREFX-associated disorder. The
CC nucleic acids can be used to express OREFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 130 AA;

Query Match 9.5%; Score 656; DB 21; Length 130;

Best Local Similarity 98.5%; Pred. No. 6.7e-32;

Matches 128; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 806 QFQPRCVSPVPVPTSRIONPVAFLSSVLPSPALPPTNAMXLPSPAPSPMSQGLAKNT 865

Db 1 QFQPRCVSPVPVPTSRIONPVAFLSSVLPSPALPPTNAMXLPSPAPSPMSQGLAKNT 60

QY 866 KSPQPVNDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISSFFQRLMNEIEFRLETPV 925

Db 61 KSPQPVNDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISSFFQRLMNEIEFRLETPV 120

QY 926 DESDDEIQHD 935

Db 121 DESHDEIQHD 130

RESULT 12

AA084351

ID AA084351 standard; Protein; 1953 AA.

XX AA084351;

DT 08-MAY-2002 (first entry)

XX Protein MYLK differentially expressed in breast cancer tissue.

XX Human; diagnosis of breast cancer; endometrial cancer; breast tumour;

XX MAI; mitotic activity index; cytostatic.

XX Homo sapiens.

XX WO200210436-A2.

PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PHD, Myers EW;
XX
XX WPI; 2001-65860/75.
DR N-PSDB; ABL02768.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclousure; SEQ ID NO 2787; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4796 AA;
XX
Query Match 7.9%; Score 545.5; DB 22; Length 4796;
Best Local Similarity 21.1%; Pred. No. 3.4e-23;
Matches 257; Conservative 170; Mismatches 484; Indels 305; Gaps 42;
XX
Oy 215 IPADTR-DNEVNHAEQOEAKREAEQAASEAGGDTTPGSSPSLYVEEPLGPPRETQ 273
Db LPSNFTGTESLQLEETHMKREE-----LVTEDEQPNPPKETE 3680
Oy 274 KLRSR-EVEPTGRVLDGIV--VGIPPPQVRWYCEGKELENSPDTHIVQAGNLSLTAIE 330
Db EIKNDLVPEGGPIHFDCEVPEVGPDPTRIEWFYNGHWATGSRVHQLNDFGFALDVIDY 3740
Oy 331 AFEEDTGRYSCFASNIYTDTSABIYIEG-----VSSSDSGDPNKEEMRIQKPNVSS 386`
Db IYARDSGEYTCRATNKGWTATTSAKVTCKGHNIVYESQLPEGMTSEKLKELEGR---- 3796
Oy 387 PPTTSVIPPVPAQHLVAOPRVATIQCCSPNTYLOGDGKPIIAAPVETKMLNLSA 446
Db -----IPEAPKVVE-----VFGPPKFTTQITSVTV 3822
Oy 447 SEGQLVVFECRY--KGAPSPKVWYREGTLIEDSPDFRILQKPRSMAPPEICTLVIAE 504
Db DEAEAVRECEQVEPKTDFSLRVWYRNGKPLPSGHRYNIFDM-----GFVSLDILY 3874
Oy 505 VPAESGCGFTCTASKNGY-----TVSS-----IAQLHV-RGNEDLSNGLSHANSTT 551
Db VYGEDSGEYVCRAINNYGDEDRATVSVCKLPTILLQNVPRG-----MKRSDALTOMEATI 3931
Oy 552 NLAATEPQSPSPHSEPPSVEOPPK-----PKLEGVLVNHNEPRSS----- 591
Db KYTSEVHLTEDDLDPDRKQPPRVTVQIKOELTITEMAVTKFECQLAPVGDPNMKVWF 3991
Oy 592 -----SRIGLRVHFNLPEDDKGS-----EASSEA 615
Db FNGKPLLHKNRFQPIYDFGYVAMNFGWVYPE-DSGEYVCRATNLYGDKETRAILKVSCKP 4050
Oy 616 GVWTRQTRPDSXQRFNGQATKTPPEPSFPKPPVLAQKPLDSTQQLQHLNQVLLBQH 675
Db GIVYDSQL-PAHQMSIDIRREAMEASQVVPVDEVPD--AKPRTRKVFVSKLEPQTVEE-- 4105
Oy 676 QLQNPPSPSPKEF-----PFKMTVLNSNAPPVATTSKKQVKAPSSQTFSLARPYF 726
Db -----GDPARFCVRVTHGPRPRVMWLNGH-----TVVHGSRYKLTNDGMFHLDPVKTR 4154
XX Homo sapiens.

Oy 727 FPST-----NTAATVAPSSSPVF-----TLSSTPQ-----T 753
Db QYDTGKVEVIARNVSGESIAATTELKVVARSDDYRNVLNKNSPRWYDYELAAAYQKEROENE 4214
Oy 754 IORTVSKESLLVSHPSVOTKSPGSLSIQNEPLPGCPTEPTPTPTPTSPGNSQPPRCVS 813
Db LKVFDERKQVLSQSSTLK--GV---EHLKPKQYKPTPTDQONVKA----- 4258
Oy 814 PIPVSPTSRIQNPVAFVLS--VLPSLPAIPPTNAXLP--RSAPSMPSQGLA----- 861
Db ---KKSDDYNNKQTLTETQELLKNETNLRDTHQVAIPGEKVVSSSQAKMAQVVEMLQ 4314
Oy 862 -KNTKSPQPVNDONIRETKNAVIRDLKKITFSVDRNQOEYKISSPEQRLMNEIERL 920
Db EKTSTTEVQAAPPKGIAQPSSESVHG-----REVHNKQ-----QQQKEIQDGL 4359
Oy 921 E--RTPVDESDETOHD-----EIPTKGCIAPIDPKRLKHFRTBGSVPTFTCKIVG 970
Db EITRKITATETTEVEHKGCTIQERVVQGPVPAKAPVFTTKIQPCRVFENEQAKFEVEFEG 4419
Oy 971 IPVPKVVYFKDGKQISKRNEHCKMRREGDGTCSLHIESTTSDDGNYTIMAANPQGRISC 1030
Db EPNPTVKWYRESFPI--ONSPDLQIHFTSGKSIILIROVFVEDSAVFCVAENRGGTAKC 4477
Oy 1031 SGHLMVQSLPIRSRLTSAGQSHRGRSRVQERDKPELQERFRPHFLQAPGDMVAHEGLRC 1090
Db SANLVVEE-----RRRACKGGIQP-----PSEVTTIQSTVTATGOLA 4514
Oy 1091 RLDCKVSGLPPELTLWLNQOPVLDPADSHKMLVRETGVHSLLDLPLTQRDAGTKYCIATN 1150
Db RFDAKVTCTRLDVLVWLNKMGKIQPSIKFKML-BEDSVHTLLIIEPFAEDSGRYEAVN 4573
Oy 1151 KTGO-----NSFSLELSVVAKEVKAPVILEKQNCVPGEGPVRLRCRVIGMPPP 1201
Db AAGEARCDGDCIVOSPSPKPEKPTTPGSEKAPHIVEQLKASQTVESGKSVIFRCRVDGRPTP 4633
Oy 1202 VFYWKK-DNETIPCTRERISMHQDITGVACLLIQPAKSKDAGWYTLTSAKNAGVISTAR 1260
Db TARMWGENFVKPSRYFQMSRQGE---YYQLVISEAPEDESTYKCAENKLGSIQTSQA 4690
Oy 1261 LDYVAQWHQIPTPMS 1276
Db LKVRPIENLADPTIT 4706
RESULT 14
ABP60991
ID ABP60991 standard; Protein; 5635 AA.
XX
XX AC ABP60991;
XX
XX DT 10-SEP-2002 (first entry)
XX
XX DE Novel human protein. SEQ ID 78.
XX
XX KW Human; cytostatic; vulnerary; antiarteriosclerotic; antiparkinsonian;
KW neurotropic; neuroprotective; immunosuppressive; haemostatic;
KW antiinflammatory; cardiact; antitumor; virucide; antithyroid;
KW cerebroprotective; anorectic; metabolic; vaccine; cancer; infection;
KW wound healing disorders; atherosclerosis; parkinson's disease;
KW Alzheimer's disease; autoimmune disorder; haematopoietic disorder;
KW inflammation; neoplastic disease; nervous system disorder;
KW cardiovascular disorders; pancreatitis; respiratory disorder;
KW hyperproliferation; systemic autoimmune disease; hyper-immunity;
KW developmental abnormality; gastrointestinal ulceration; neuropathy;
KW haematological disease; metabolic disease; sperm dysfunction;
KW thyroid disorder; hypothyroidism; brain damage; colitis;
KW cone photo- transduction deficiency; neurological disease; stroke;
KW angiogenesis; ovulation disorder; spinal cord; thyroid gland; heart;
KW trachea; thymus; lymph node; muscular system; obesity; anorexia;
KW growth abnormality; precocious puberty.
XX
XX OS Homo sapiens.

QY 1313 SVV 1315
: |
Db 2927 NNV 2929

Search completed: November 30, 2002, 19:04:50
Job time : 62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 19:03:55 ; Search time 19 Seconds
(without alignments)
2044.119 Million cell updates/sec

Title: US-09-818-990B-2
Perfect score: 6890
Sequence: 1 MODSIEASTSIQLRESY.....MESTMYSCSRVSDEL 1320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	729	10.6	498	US-09-354-151-2	Sequence 2, Appli
2	335	4.9	277	US-09-354-151-3	Sequence 3, Appli
3	321.5	4.7	1297	US-09-540-245A-17	Sequence 17, Appl
4	318.5	4.6	1395	US-09-540-245A-15	Sequence 15, Appl
5	314	4.6	1651	US-09-540-245A-18	Sequence 18, Appl
6	311	4.5	601	US-08-795-868-16	Sequence 16, Appl
7	311	4.5	602	US-09-303-069-16	Sequence 16, Appl
8	311	4.5	602	US-09-134-250-16	Sequence 16, Appl
9	303	4.4	661	US-08-795-868-14	Sequence 14, Appl
10	303	4.4	661	US-09-303-069-14	Sequence 14, Appl
11	303	4.4	661	US-09-134-250-14	Sequence 14, Appl
12	300	4.4	1447	US-09-041-886-25	Sequence 25, Appl
13	300	4.4	1447	PCR-US94-05277-2	Sequence 2, Appli
14	267.5	3.9	607	US-08-752-307B-12	Sequence 12, Appl
15	267.5	3.9	607	US-09-707-802-12	Sequence 12, Appl
16	267.5	3.9	607	US-09-991-326-12	Sequence 12, Appl
17	263.5	3.8	612	US-08-752-307B-11	Sequence 11, Appl
18	263.5	3.8	612	US-09-707-802-11	Sequence 11, Appl
19	263.5	3.8	612	US-09-991-326-11	Sequence 11, Appl
20	263.5	3.8	1268	US-08-506-296B-28	Sequence 28, Appl
21	260.5	3.8	615	US-08-752-307B-9	Sequence 9, Appli
22	260.5	3.8	615	US-09-707-802-9	Sequence 9, Appli
23	260.5	3.8	615	US-09-991-326-9	Sequence 9, Appli
24	260.5	3.8	1381	US-09-540-245A-16	Sequence 16, Appl
25	257	3.7	380	US-09-877-730-4	Sequence 4, Appli
26	257	3.7	904	US-09-877-730-6	Sequence 6, Appli
27	257	3.7	985	US-09-877-730-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-354-151-2
; Sequence 2, Application US/09354151
; Patent No. 6458929
; GENERAL INFORMATION:
; APPLICANT: CARPEN, Olli
; APPLICANT: GRONHOLM, Mikaela
; APPLICANT: HEISKA, Leena
; APPLICANT: MYKKANEN, Olli-Matti
; APPLICANT: SALMIKANGAS, Paula
; TITLE OF INVENTION: Myocilin, A No. 6458929e1 Actin-Organizing Protein
; FILE REFERENCE: 0933-0142P
; CURRENT APPLICATION NUMBER: US/09/354,151
; CURRENT FILING DATE: 1999-07-16
; EARLIER APPLICATION NUMBER: 60/093,169
; EARLIER FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-354-151-2

Query Match	10.6%	Score 729;	DB 4;	Length 498;
Best Local Similarity	29.3%;	Pred. No. 4.5e-41;		
Matches 185;	Conservative 101;	Mismatches 182;	Indels 164;	Gaps 16;
QY	718	FSLARPKYFPSTNTTAAVTAPSSPVFTLSSTPQTQRTV-----SKESLLV	765	Sequence 2, Appli
Db	2	FNVERPKHFIQSQNPCGSRLLQPGPETSFSFS--QTKQSSIIIQPROCTEQRFSASSTLS	59	Sequence 8, Appli
QY	766	SHPSVQTKSPGGLSIONEPLPGPTEPTPPPTFTSIFSGNQFQRCVSPVPSPTSRI--	823	Sequence 52, Appli
Db	60	SH-----ITMSSSAFPASPOQ-----HAGS-----NFCQRTVT	87	Sequence 5, Appli
QY	824	---QNPVAFSLSSVLPSLPAIPPTNAMXLPSPASMPQGLAKK--NTKSPQPVN-----	872	Sequence 2, Appli
Db	88	TYNGSPASFLSSILPSQ---PDYNSSKIPSAMDSNYQSSAGQFINAKPSQTANAKPIPR	144	Sequence 14, Appl
QY	873	--DDNIRETKNAVTRDGLKKTITFSDVRPNQOEYKISSFEQRLMNEIEFRLERTVPVDSDD	930	Sequence 14, Appl
Db	145	TPDHEIQGSKREALIQDLERLKKCD-----TLLHNGNQ	177	Sequence 14, Appl
QY	931	EIOHDEIPTCKIAPIDFKRLKHEFVTEGSPVTCTKIVGPVKVYWFKDGKISKRNE	990	Sequence 104, App
Db	178	RLTYEERKMARLLGP-----ONAAVFAQ-----	202	Sequence 62, Appl
QY	991	HCKMRREGDGTCSLHISTTDDDDGNTYIMAA NPQGRITSCGHLMVQSLPIRSRLTSAGQ	1050	Sequence 13, Appl


```
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-245A-15

Query Match      4.6%; Score 318.5; DB 4; Length 1395;
Best Local Similarity 26.3%; Pred. No. 7e-13;
Matches 108; Conservative 50; Mismatches 155; Indels 97; Gaps 14;

QY 956 VTEGSPVTFCKIVGIPVYVFKGQKQIKRNECHKMRREGDGTCSLH--IESITSDD 1013
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 VKKNEPATLNCKVEGKPEPTIEWFKDGEPVSTNEKKSHRVQFKDGALEFFYRTWQKKEQD 126
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1014 DGNYYTMAANPOGRISCSGHLVQSLPIRSRLTSAGOSHRGRSVROBRDKEPQERFP 1073
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 GGEYWCYAKNRVGO-AVSRH-----ASQIAVLRLD 155
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1074 HFLQAPGDVMAHGRGLRCLDC-KVSGLPPELTLWLANGQVLPD-----ASHKMLVRE 1125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156 DFRVEPKDTRVAKGETALLECGPPKGIPEPTLIWIKDGVPLDLDKAMSGFASRRVRIVD 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1126 TGVHSLLDLPTORDAGTYKCIATNKTGQNSFSLSVAKVKKAPVILEKLNQCGVPE 1185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 GG--NLLISNVEPIDEGNYKICIAQLNLTGRTRESSYAKLIV--QVK--PYFMKEPKDQWMLY 268
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1186 GHPVRLRCRVIGMPVPVYVYKKNONETIPCTRERISMHDQTTGYACLLIQAPAKKSDAGWYT 1245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 GQATFHCSVGGDPPPKVLMKKEBGNIPVSRARI-LHDEKS-----LEISNITPDEGTIV 323
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1246 LSAKNEAGIVSCTARLDIYAQWHQIIPPPMSVRPSGRYG-----YSCSRSVVES 1317
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 CEAHNNVGQISARASLIVHA-----PPNFKRPSNKKVGLNGVQLPCMASGNPPPSVF 377
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1286 -----SLTSKGLDIFSAFSSMESTWV-----YSCSRSVVES 1317
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 378 WTKEGVSTLMPNPNSSHGQRVAAADGTLOITDVRQDEGYVCSAFSVYDS 427
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-09-540-245A-18
; Sequence 16, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT FILING DATE: 2000-03-31
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1651
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-18

Query Match      4.6%; Score 314; DB 4; Length 1651;
Best Local Similarity 22.6%; Pred. No. 1.8e-12;
Matches 134; Conservative 68; Mismatches 211; Indels 180; Gaps 21;

QY 764 LVSHPS---VQTKSPGGLSIONEPLPPGPT-----EPTPPPTFSIPSGNQF 807
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 IVEHPSDLIVSKGEPATLNCKAERGPT-PTIEWKGERVETDKDDPRSHRMLPLPSGLF 128
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 808 QPRCVSIPVPSPTSRIONQPNVAFSLVSLFAIPPTNAMLPRAPSMPQSGL---AKKN 864
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 129 FLRIV-----HGRKSRPDEGVYVCVARN 151
QY 865 TKSPQPVNDNIRETKNAVRLDLGKKITFSDVRPNQOEYKISSFEQRLMN----- 914
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 152 YLGEAVSHNASL---EVALIRD-----DFRONPSDMVAVGEPVAMECQPPRGHPEP 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 915 EIEFRLERTPVDESDDDEIQHDEIPTGK-----CIA----- 944
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201 TISWKKDGSPLDDKDERI---TIRGGKLMITYTRKSDAGKVVCVGTNMVGERESEVAELT 257
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 945 ----PIFDRLKLRHVRTEGSPVTFCKIVGIPVYVFKDQKQISK-----RNEHCKMR 995
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 258 VLERPSVKKRPSNLAVTVDDSAEFKCEARGDPVPTVRWKDDGELPKSRYEIRDDH---- 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 996 REGDGTCSLHESITSDDDGNYTITMAANPOGRISCSGHLVQSLPIRSRLTSAGQSHRGR 1055
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 314 -----TLKIRKVTAGDMGSYTCVAENMVGKAASATLTIVQ----- 348
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1056 SRVQERKEPLQERFFRPHFLQAPGDVMAHGRGLRCLDCVSGLPPELTLWLANGQVPL- 1114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 349 -----EP-----PHFVVKPRDOVVALGRVTVTFOCEATGNPQPAIFWRREGSQNLL 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1115 ----PDASHKMLVRETGVHSLLDLPTORDAGTYKCIATNKTGQ--NSFSLELSVVAKE 1167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 FSYQPPQSSSRFSYSQTG--DLTITNVQSRSDVGYIQCITLNVAGSIITKAYLEVTDVIA 451
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1168 VKKAPVILEKLNQCGVPEGHPVRLRCRVIGMPVPVYVYKKNONETIPCTRERISMHDQTTG 1227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 452 -RPPVIRQGPVNTAVDGTFLVSCVATGSPVPTILWRKDGVLVSTQDSRIKOLEN--- 507
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1228 YACLLIQAPAKKSDAGWYTLSAKNEAGIVSCTARLDIYAQWHQIIPPPMSVRPS 1280
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 -GVLIQIRYAKLGDGRYTCIATSPSGEATWSAYIEV-QEFGVGVQPPRPPTDPN 558
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-08-795-868-16
; Sequence 16, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; TITLE OF INVENTION: AND STRIATED-MUSCLE CELL ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
```


Db 134 EYRLERSPVDESGDEVQYCDVPVENGMAPFFEMKLVKHYKIFEGMPVTFCTRVAGNPKPKI 193
Qy 977 YWFKDGKIOISKRNECHKRRREGDGTCSLHISTTSDDDGNVTIMAAANPQGRISCSCGLMV 1036
Db 194 YWFKDGKQLSPKSDHYTITQDLDDGTCSLHTASTLDDDDGNVTIMAAANPQGRISCGRLMV 253
Qy 1037 QSLPRLSRITSAGQSH-----RGRSRVQERDKQLQRFRRPFLQAPGDMVAHEGRLC 1090
Db 254 QAVNQRGSRPSRPSGPHVRRPRSRSDGDNENIQRFRFPFLQAPGLTVOEGKLC 313
Qy 1091 RLDCVSGLPPELTWLLNGQVPLPDASHKMLVRETGVHSLIDPLTQORDAGTYKCIATN 1150
Db 314 RMDCKVSGLPTPLDLSQWLDGKVPVPSASHKMLVRENGVHSLIIEPVTSRDAGITTCIATN 373
Qy 1151 KTQNSFSLELSVVAKEVKKAPVILEKLQNCVPGCHPVRLRCRVIGMPPPVFYWKKDNE 1210
Db 374 RAGQNSFSLELVAAKEHPPVFIEKIQNTGVADGYPVRLRCRVIGLVPVPPQIFWKKENE 433
Qy 1211 TPTCTRERISMHQDTGYACLLIQPAKSDAGWYTLISAKNEAGIVSCTARLDIYAOWHH- 1269
Db 434 SLTHSTDVSMHQDNHGVICLLIQGATKEDAGWYTLISAKNEAGIVSCTARLDVYTQWQQ 493
Qy 1270 -QIPPPMSVRPSGSGYSLTSKGLDIFSASFMSMSTWYSCSSRVESDEL 1320
Db 494 QSSTKPKKVRPSASRYAALSQGLDIAKAFQ- EANPSKLTINTALVESEDL 544

RESULT 2

US-09-764-853-638
; Sequence 638, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZJ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 638
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-853-638

Query Match 20.2%; Score 1394; DB 10; Length 264;
Best Local Similarity 99.6%; Pred. No. 9,8e-60;
Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1059 QERKEPLQERFFRPHFLQAPGDMVAHEGRLCRLDCKVSGLPPPELTWLLNGQVPLPDAS 1118
Db 3 RERDKEPLQERFFRPHFLQAPGDMVAHEGRLCRLDCKVSGLPPPELTWLLNGQVPLPDAS 62
Qy 1119 HKMLVRETGVHSLIDPLTQORDAGTYKCIATNKTQNSFSLELSVVAKEVKKAPVILEKL 1178
Db 63 HKMLVRETGVHSLIDPLTQORDAGTYKCIATNKTQNSFSLELSVVAKEVKKAPVILEKL 122
Qy 1179 QNCVGPEGHPVRLRCRVIGMPPPVFYWKKDNETIPTCTRERISMHQDTTGYACLLIQPAK 1238
Db 123 QNCVGPEGHPVRLRCRVIGMPPPVFYWKKDNETIPTCTRERISMHQDTTGYACLLIQPAK 182
Qy 1239 SDAGWYTLISAKNEAGIVSCTARLDIYAQWHHQIIPPMNSVRPSGSGRYSLTSKGLDIFSAP 1298
Db 183 SDAGWYTLISAKNEAGIVSCTARLDIYAQWHHQIIPPMNSVRPSGSGRYSLTSKGLDIFSAP 242
Qy 1299 SSMESTWYSCSSRVESDEL 1320
Db 243 SSMESTWYSCSSRVESDEL 264

RESULT 3

Query Match 7.0%; Score 485; DB 9; Length 7968;
Best Local Similarity 19.0%; Pred. No. 4.3e-15;
Matches 305; Conservative 186; Mismatches 527; Indels 590; Gaps 55;

US-09-764-853-817
; Sequence 817, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZJ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 817
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (121)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (139)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-817

Query Match 10.5%; Score 724.5; DB 10; Length 159;
Best Local Similarity 89.6%; Pred. No. 2.6e-28;
Matches 136; Conservative 3; Mismatches 10; Indels 3; Gaps 2;
Qy 1142 GTYKCIATNKTQNSFSLELSVVAKEVKKAPVILEKLQNCVPGCHPVRLRCRVIGMPP 1201
Db 8 GTYKCIATNKTQNSFSLELSVVAKEVKKAPVILEKLQNCVPGCHPVRLRCRVIGMPP 67
Qy 1202 VFYWKKDNETIPTCTRERISMHQDTTGYACLLIQPAKSDAGWYTLISAKNEAGIVSCTARL 1261
Db 68 VFYWKKDNETIPTCTRERISMHQDTTGYACLLIQPAKSDAGWYTLISAKNEAGIXSCTARL 127
Qy 1262 DIYAQWHHQIIPPP-MSVRPSGSGRYSLTSKGLDI 1294
Db 128 DIYAQWHHQIIPPCLPDPAQXSXRISN--SKGLDI 159

RESULT 4

US-10-077-130-5
; Sequence 5, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: WPI2001-047P1RCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7968
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-5

Query Match 7.0%; Score 485; DB 9; Length 7968;
Best Local Similarity 19.0%; Pred. No. 4.3e-15;
Matches 305; Conservative 186; Mismatches 527; Indels 590; Gaps 55;

FILE REFERENCE: 540579-2007.2
CURRENT APPLICATION NUMBER: US/09/905.129
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/802.318
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/207.821
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/084.944
PRIOR FILING DATE: 1998-05-11
PRIOR APPLICATION NUMBER: 60/085.673
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent in version 3.0
SEQ ID NO 13
LENGTH: 2597
TYPE: PRT
ORGANISM: Rattus sp.
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-13

Query Match 6.0%; Score 416.5; DB 10; Length 2597;
Best Local Similarity 19.1%; Pred. No. 2.2e-12;
Matches 354; Conservative 216; Mismatches 552; Indels 733; Gaps 83;

```
QY 2 QDDSTEASTSISQLR---ESVLAETRIRGNR---SRAEPSSNPCHFGSPSGNAEGG 55
DB 296 QBDNGSASTSPQDFTEPPGSLNMTXXGNKADMYCQIKPSRT-----SPTAFTE--- 347
QY 56 QDDLDLSAFLSQBELDESINLARLAINYDPLEKADETQARKRLSPDQMKHSPNLSFEP 115
DB 348 -ENDYIMLNASF-----SINLV-CSVDYNHIOVPWQALLALYSDSLILERKQLT--- 395
QY 116 NFOQDNRPSPSKESQEAQRQVCST-----QSKVF--LNAKADFIEL 161
DB 396 -----ETPSLSRYKQVALRPEDIFTSIEADVRADPFWFQOEKIVLQLNRTATTLSTL 448
QY 162 SSLKSHSKRIRPRACKNHKSKLESQNKVMOENSSFSDLSEERERS-----SVP 212
DB 449 QIQFSTDAQIAL-PRA-EMRAERLKWTLWMNN-----PKLERITVLVGGTIALSCP 498
QY 213 -----IPIPADTRONEV-----NHAEQQEAKRREA---EQRASEAAG 247
DB 499 KGKGDSPHLEMLADGSKVRAPYVSEDGRILLDKNGKLEQWADSFDAGLYHCISTNDAD 558
QY 248 GTTTPGSSPSLYE-----EPLGQPPRTQKLRSEVPEGRVQDLCIVGIPPPQVRWY 303
DB 559 ADV-----LTYRITVPEYGESTHDSGV--QHTVVTGTETLDLPCLSTGVPDASISWI 608
QY 304 CEGKELENSP--DIHIVQAGNLHSLTIAEAFEEDGRYSCFASNIYGTGSTSAEYI--- 358
DB 609 LPGNTVFSQPSDRILLNGLTRILQVT---PKDGHYQOCVAANPSGADFSFKVSVQKK 665
QY 359 -EGVSSDSE-----GDPNKEEMRIQKPNREVSPPTTSVAPPVAP----- 399
DB 666 GQRMVYEHDRAGSGGLGEPNPSVSLKQASLKLKSALATSEAGKQVSGVHRKNNKRDLI 725
QY 400 -----QAHLVAQPRVATTQCCQ--SPTNYLOGLDGKPI-IAAP 435
DB 726 HRRRGDSTLRRFRHRRLQPLSARIDPQRAWALKEKAKNSVPKKQENTTVKVPPLAVP 785
QY 436 VFTKMLQNLSA---EGOLVVEFCRVKGP--SPKVE-----WYREGTLI--- 475
DB 786 LVELTDEKDKASGMTPPDEEFVWLTKASGVGPRPTADSGPVNIGFMTSTASGTEFVSTV 845
QY 476 -----EDSPDR-----ILQKPRSMAPPEICT-----LVIAEFAE--DS--- 510
DB 846 NPQTLQSEHLPDFKLFVNTNGTAVTKSMNPSIAKIEDTTQNPIIIFPSVAEIRDSQA 905
QY 511 -----GCTCTAS-----NKYGTVSSTAQLHVRNE 536
```

```
DB 906 GRASSQSAHPVTGGMATYGHNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSS 965
QY 537 DLSNNG-----SLHSANSTTNLAAIEQP-----Spph 564
DB 966 DISHTTADPSFSSHPSGSHTTASSLPHIPRNNNTGNFPLSRHLGRERTIWSGRVKNPH 1025
QY 565 SEP-----PSVEQP----- 573
DB 1026 RTPVLRHRHRTVRPAIKGPANKNVQVPATEYPMCHTCPSAEGLTVATAALSVPSSH 1085
QY 574 ---PKPKLEGVLV----- 594
DB 1086 SALPKTNVGVIAEESTTVVKKPLLFKDKQNDVIEIITTTKYSGGESNHLVITEASMT 1145
QY 595 GLRVHFNL---PEDDKGSEASSEAGVVTROTDPDSXQERENGQATKTPEPSFPVKBP- 650
DB 1146 SAPTSVSLGKSPVDSNG-----HLSMPTGIQTIGKDSV-----TTPLES-PLSTPSI 1191
QY 651 PVLAKPKLDSTLOQL--HNO---VLLLEOHL--QNPPSPSPK-----EFPFXM 692
DB 1192 PTSTKFSKRKTPLHQIFVNNQKKEGMLKNPYQFGLQKNPAKLPKIAPLLPTGQSSPDS 1251
QY 693 TVLSNAPPAVTT-----XQVKAPSSQTFSLARPKYFFPST-----NTT 733
DB 1252 TLTLSPPALSTTMAATQNKGTVEVSGARLSAGKKQPFNTSSP--VLPSTISKRSNTL 1309
QY 734 -AATVAPS-SSPVTLSSTPQTIQRTVSKES----- 762
DB 1310 NFLSTETPTVTPATASVIMSETQRTSRKEAKDOIKGPRKNRNNANTTPROVSGYSAYS 1369
QY 763 -----LLVSHPSVQ----- 772
DB 1370 ALTTADTFLAFSHSPRODGGNVSAVYHSTTSLAITELFEKYTQTLGNTTALETLLS 1429
QY 773 KSPGGLSTON-----EPL-----PPGTEPTPPPTFTSIPSGNQ-----FQPRCVSPPVS 818
DB 1430 KSQESTTVKRASDTPPPPLSSGAPVPT-PSPPPTKGVVDSKVTSFAFOTNRVVTIY 1488
QY 819 PTSR---IQNPVA-----FLSSVLPSPALPPTNAMXLPSPAPSM 855
DB 1489 ESSRHNTDLOQPSAEASPNELIITGTDSPSNLFPSTVPAIRVVDKPKQNSKWP----- 1542
QY 856 PSQGLAKNKTSPQPVNDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISFQRLMNE 915
DB 1543 -----SPWP-----EHYQLKSYSETIEKG 1562
QY 916 IEFRLRTPVDESDEIHOHEIPTGKCTA-----PIFDKR----- 950
DB 1563 KRPAVMSNP-----HLSLPEASTHASHWNTQKHAESVFDKKPKQCNQNTSKHLPYVS 1613
QY 951 -----LKHFRTVEGSPVFTT-----CKIVGIPVPKVYWFK--DGKQISKRNEHCK 993
DB 1614 LPKTLLKPRIGGKAASFTVPANSDFVLPCEAVGDPLPIHWTVRVSGXELISQGTQSR 1673
QY 994 MRREGDGTCSLHIESTTSDDDGNVTIMAAANPQGRISCSGHLMVQSLPRTSLTSAGQSHR 1053
DB 1674 FHVLNPGT--LSIQKVSIOQDRGVQLCSAFNPLGVDPHFHVSLSVVEYP----- 1718
QY 1054 GRSRVOERDKPELQERFPRPHFLQAPGDMAHEGLRCLDCKVSCGLPPELTWLLNGQPV 1113
DB 1719 --ARTLDHRVK-----EITVHFGSTVEKCRVEGMPRPTVSWILANQTV 1760
QY 1114 LPD---ASHKMLVRETGVHSLLLIDPLTORDAGTYKCIATNKTGNSFSLSVVAKEVKK 1170
DB 1761 VSETAKSGKVVWTPDG--TLIYNLSLYDRGFYKCVASNPSGQDLSLLVKIQVIT--- 1814
QY 1171 APVILEKIQNC--GYPEGHVRLCEVRIGMPPPVFYWKNDNET-----IPCTRERISMHD 1224
DB 1815 PPVITEQRQAIQVG-LGGSLLKLPCTAKGTQPSVHWVLYDGTGELKQLQLTHSRFFLYPN 1873
QY 1225 TTGYACLLIQAKKSDAGWYTLAKNEAG----- 1253
DB 1874 GT----LYIRIAPSVRGTYECIATSSSGSERRVVILTVERGETIPRIETASQKTEVNL 1929
```



```

QY 1263 IYAQWHQHOIPPPMVSVPSCS 1282
   11 11 11 11 11 11 11 11
Db 2487 IVA-----YPPRITNRPPRS 2501

RESULT 14
US-09-991-630-14
; Sequence 14, Application US/09991630
; Patent No. US2020151514A1

```

RESULT 14

```

US-09-991-630-14
; Sequence 14, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991,630
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/7905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 2586
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-991-630-14

```

```

Query Match      6.08; Score 410; DB 10; Length 2586;
Best Local Similarity 19.4%; Pred. No. 4.4e-12;
Matches 283; Conservative 187; Mismatches 522; Indels 468; Gaps 59;

QY      8 ASTISOLLRESYLAET---RRHGNNERSRAPSSNP---CHFGPSGAAGGGGDDLPD 62
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1325 ASVITYETQTRSRAQTIOREQEPQKNRNTDPNISPDQSSGFTPTAMT-----PP 1375

QY      63 LSAFLSQEELDESYNLARLAINYPLEKADETQARKRLSPDMQKSPNLSPEPNCQNP 122
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1376 ALATFHPSPENTTGISSTFISHSTNLJTDVIELAQASTQTLKST--IASETTLSSKSH 1433

QY      123 RSPSTSSKESPOEAKRPQVCSETQS-----KKVFLNKA-----ADPTEELSSLFKSHS 169
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1434 QSTTTTKAS-LDTPPIPFELSSSATLMPVIPSPPTQRAVTDTRGDSHFRLMTNTVVKLHE 1492

QY      170 SKR-----IPRACKN---HKSLESQNKVMQENS-----SFSDL 202
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1493 SSRINLQMPSQLBPLTSTSNLHJSTPMPAULTTVKSONSKLTPSPWAEYOFWHKPYSDI 1552

QY      203 SERRERSVPIPADTRDNEVNHAE-----QOEAKRREAE-QAASEAAGGDTTGPCSSPS 257
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1553 AEKGGKKEV--SMLATGLSEATTLVSDWDQGNKTKKSDFKKPVQEATTKLLPFDLSL 1610

QY      258 SLYEEPL---GQPRFTQKLRSEVPGETRVQOLDICIVVGIPPPQVRW-YEGKELNSP 313
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1611 RYIEFKPRIVGGKASET-----IPANSDAFLPCEAVGNPLPTIHTRVSGLDLSRGN 1663

QY      314 DIHIVQAGNLHSLTAAEAFEDTGRYSCFASNIYGTDTSAEIXIEGVSSDSEGDPNKE 373
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1664 QNSRVQVLNPGTLSIORVEIQDROGYLCSANLFGTDHLHVLTVS-----1708

QY      374 EMNRIQKPNEVSSPPTTSVAPPVQAQHLVAQPRVATIOCQCSPTNYLOGLDGKPIIA 433
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1709 -----VSYPPE-----1714

QY      434 APVETKMLQNLASSEGOLVVEECRVKGAQSPKVEW-YREGTLIEDSPDFRILQKPRSWA 492
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db 1715 --ILERTKETIVHSGSTVELKCRAGRPSPVTVMILANQTVWSES-----SQG 1761

QY      493 EPEEIC----TLVIAEVAEFDSCGFTCTASNKYGTIVSSIAQLHVRGNDLSNNGLSHSAN 548
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||

```

Db	1762	SROAVVTVDGTLVHLNLSYIDRGFYKCVASNPGGQDLSLLVQIV-----	1808
Qy	549	STTNLAAIEPQPSP-----HSEPPSVQEPQPKLEGVLVHNEPR--	589
Db	1806	-----IAA-----PPVILEQRQVIVGTWGESLKLPCTAKGTQPPSVYVWLSGTEVKPL	1855
Qy	590	-----SSRIGRLRVHNLPEDDKGE--ASSEAG-----VVTRTRTRDPSXOER	631
Db	1856	QFTNSKLFLESNGTLIR--NLASSDRGYTECIATSSGTERRRYVMLTMEERVTS--PR	1910
Qy	632	FNGOATKTPSPVPKPEPPVLAKPKLDSTQLQHLNVLLEQHLQNLQPPSPSKPEFPFX	691
Db	1911	IEAASQARTEVNF-----GDKLLNCSATGPEKP-----	1939
Qy	692	MTVLNSNAPPAVTTXKQV--KAPSSQTFSLARPKYFFPSTNTTAATVAPSSSPVFTLSS	749
Db	1940	-----QIMWRLPSKAAVDQGSWIHYVNGSLFGVSTEKDSGYL--	1979
Qy	750	TPQTIQRTVSKESLLVSHPSVQTKSGGLSIQNEPLPGPTEPTPTPTTFPSIPSGNOFQP	809
Db	1980	--CVARNKMGDDLILMHVSLRLK-PAKIDHKQY-----PRKQVLHGCKQFQV	2022
Qy	810	RCVSPIPVSTFIQNPVAFLLSVLPSLPAIPPTNAMXLPSPASPMSPQGLAKNTKXSPQ	869
Db	2023	DKRAS-----GSPVPEISWSLPGDTMI--NNAMQA-----	2050
Qy	870	PVNDNIRETKNAVIRDLGKLFIFSDVRPNQOEYKISSEQRNLMEIEFLERTPVDESQ	929
Db	2051	---DDSGHRTYRTYLFNNG--TLFYKNQGV-ABEGDYTCVAQNTLCKDEMKVHLVITAAP	2105
Qy	930	DEIGHDEIPTKCIAPIFDKRLKHFRVTBEGSPVTTCKTIGVIPVKVYV-FDKGQKISKR	988
Db	2106	RIRQSNK--INK-----RIKAGDTAVLDCVETGDPKPKIEFWLLPSNDMISFS	2150
Qy	989	NEHCKMRREGDGTCSLHISTSDDCGNTYIMAAQNPQGRISCSGHLMVQSLP-----	1040
Db	2151	IDRYTFHANG-----SLTINKVKLLDSGEYVCVARNPSGDDTKMYKLDVYVSKPPLINGLYT	2206
Qy	1041	-----IR-----SRLT-----	1046
Db	2207	NRTVIKATAVRHSKKHFDCAEGTSPSEVYMWIMPONIPLTAPYQSRITVHKNGTLEIRN	2266
Qy	1047	-----SAGQSHRGRS--RVQBRDKEPQOERFFRPHFLQAPDMVAHEGRLCRDLC	1094
Db	2267	VRLSDSADFICVARNEGGEVLVVQLEVLMLRRPTFRNPFE--KIVAQLGKXSTALNC	2323
Qy	1095	KVSGLPPPELTWLL-----NCQVPLPDASHKMLVRETVGHSLLIDPLTQBDAGTYKCI	1147
Db	2324	SVDNPPPEIWIPLNPTRESNG-----PQSYQYLIASNG--SFIIKSTRTEDAGKYRCA	2376
Qy	1148	ATNKGQNSFSLELSVVAKEVKKAPVILEKLQNC--GVPEGHPVRLCRCVIGMPVPFYV	1205
Db	2377	ARNKVGY-----IE-KVILEIGQKPVILTYAGCTVKGI--SGESLSLHCVSQGIKPNIKW	2430
Qy	1206	KKDNETI---PCTRERISMHQDPTYGACLLIOPAKKSADAGWTTLSAKNEAGIVSCTARLD	1262
Db	2431	TMPSGVYVDRPQINGKYLHDNGT----LVKEATAYDRGNVICAQNSVGHITLTVPM	2486
Qy	1263	IYAQWHHQIIPPPMSVRPSGS	1382
Db	2487	IVA-----YPPRTNRRPPRS	2501

RESULT 15

```

US-905-129-16
; Sequence 16, Application US/09905129
; Patent No. US2002037705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905.129

```


Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1666	24.2		772	2	T13078	KIAA0932 protein -
2	871.5	12.6		4162	2	T42633	connectin/titin -
3	801.5	11.6		26926	1	I38344	titin, cardiac musc
4	608	8.8		6842	2	S68235	protein UNC-89 - C
5	556.5	8.1		1906	1	T29737	myosin-light-chain
6	531.5	7.7		7962	2	I38346	elastic titin - hu
7	520.5	7.6		811	2	PN0689	connectin 1 - chic
8	513	7.4		2783	2	T34416	hypothetical prote
9	502	7.3		2541	2	T29340	hypothetical prote
10	468.5	6.8		5175	2	T20992	hypothetical prote
11	468.5	6.8		5198	2	T43290	hemiscitin precurs
12	429.5	6.2		1021	2	T43634	connectin/titin -
13	402	5.8		6831	2	A88852	protein unc-22 (im
14	402	5.8		6839	2	S57242	twitchin (similar)
15	402	5.8		7160	2	T27935	hypothetical prote
16	397.5	5.8		4391	2	A38096	perlecan precursor
17	375.5	5.4		3707	2	S18252	heparan sulfate pr
18	374	5.4		1323	2	PN0568	connectin 3B - chi
19	364	5.3		1176	2	JN0583	myosin-light-chain
20	352.5	5.1		1147	2	A59307	myosin-light-chain
21	351.5	5.1		6658	2	T13931	projectin - fruit
22	329	4.8		1273	2	T42405	sax-3 protein - Ca
23	322	4.7		3375	2	T19821	hypothetical prote
24	321.5	4.7		423	2	T29549	hypothetical prote
25	319	4.6		1612	2	T30805	dufil protein - mo
26	317	4.6		1465	2	S43529	165K protein, ske
27	316.5	4.6		1344	2	T14316	rig-1 protein - mo
28	315.5	4.6		1651	2	T14160	transmembrane rece
29	315	4.6		1896	2	T08851	Dow syndrome cell

```
Db 275 SRTASDEIQTKDAVIDGLERKLKFKEDLLNNGQPRLTYYERMARRLLGADSATVF 334
Qy 900 -----QOEYKISSFEORLMEIEFRLERTPVDESDDIEIOHDEIPTCKIAPIDFK 949
Db 335 NIQPEETANQEVKVSCEQRLISEIERYLERSFPVDESQVIGDVPVENGMAPFEM 394
Qy 950 RLKHFRTVEGSPVFTCKIVGIPKVVYFKDGKQISKRNEHCKMRREGDGTCSLHTEST 1009
Db 395 KLKHYKIFEGMPVFTCRVAGNPKKIYFKDGKQISPKSDHYTIQRLDGTCSLHTTAS 454
Qy 1010 TSDDGNTYIMAAPOGRISCSGHLMVQSLPIRSLRISAGOSH-----RGRSRVQERDK 1063
Db 455 TLDDGNTYIMAAPOGRISCTGLMWQAVNQGRSPRSPSCHPRRPRSRSDSGDEN 514
Qy 1064 EPLQERFRPHFLOAGDWAHEGLRCLCKVSLGSLPPELTLLNGQVPLPDASHKMLV 1123
Db 515 EPIQERFRPHFLOAGDGLTVQEGKLCMDCKVSLGSLPFDLSWLDGKVPKRPDSAHKMLV 574
Qy 1124 RETGVHSLIDPLTORAGTYKCIATNKTKQNSFSLSVVAKEVKKAPVILEKLQNGV 1183
Db 575 RENGVSIIIEPVTSROAGIYTCIATNRAGQNSFSELVVAKEAHKPPVPIEKLQNTGV 634
Qy 1184 PEGHPVRLCVRIGMPVPVFTWKNDNETIPCTRERISHQDQTTGYACLLIQAPAKSDAGW 1243
Db 635 ADGYPVRLCVRIGVPPQIPFKKKNESLHSTDRVSMHQDNHGYICLLIQGATKEDAGW 694
Qy 1244 YTLAKNEAGVCTARLDIYAOMHH--QIPPPMSVRSGSRYSGLSKGLDIFSASSM 1301
Db 695 YTVSAKNEAGVCTARLDVYTNQHQSSQSTKPKKVRPSASRYAALSDDQGLDIKAAFP- 753
Qy 1302 ESTMVYSCSSRSVSEDEL 1320
Db 754 EANPSHLTLNLTAVEDEL 772

RESULT 2
T42633
connectin/titin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42633
R:ajima, H.; Ohtsuka, H.; Kawamura, Y.; Kume, H.; Murayama, T.; Abe, H.; Kimura, S.; Ma
Biochem. Biophys. Res. Commun. 223, 160-164, 1996
A:Title: A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle connectin/titin re
A:Reference number: 222221; MUID:96254045; PMID:8660363
A:Accession: T42633
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4162 <YAJ>
A:Cross-references: EMBL:D83390; NID:g1513029; PIDN:BAAL1908.1; PID:g1513030
A:Experimental source: breast muscle
C:Keywords: skeletal muscle

Query Match
Best Local Similarity 23.5%; Score 871.5; DB 2; Length 4162;
Matches 327; Conservative 193; Mismatches 507; Indels 367; Gaps 43;

Qy 209 SSVPIPIPADRDNEVNH-----LEQOEAKREAEQAASEAAG-----DITPGSSPS 257
Db 30 SGFPVPEVSWYRGQVLSAATLPQGVISFSDGRAKLVPISVTEANSRYTITQATNGSGQA 89
Qy 258 SLYVEEPL---GQPRFTQKLRSREVPGETRVQDLCIVVGPVPPQVRWYCEGKELENSPD 314
Db 90 TSTAELLVTAGTAPNFSORLOSMTARQGSQVRLDVRVTGIPTPVKPYRGVGEIQSPD 149
Qy 315 IHIVQANLHSLTAEAFEDTGRYSFASNIYGTDSABIIYEGVSSSDSEGDPNKE- 373
Db 150 FOILQEGDLSLIAEAYPEDSGTYSYNATNNVGRATSTABELLIQ-----EEEAAPAKT 205
Qy 374 -----ENNRIOK-----PNEVSSPPTS 391
Db 206 KTIYSTAQISQTOARIEKKIETHFDARSLTSEVMWIEGAAAOQLPHKAPRMPPTSK 265
```

```
Qy 392 AVIPPAVPAQHLVAQPRVATIOQCQSPNTYNOGLDGKPIIAADPVFTKMLQNLSASEGOL 451
Db 266 SPTPEVITAKAQAQMARQSPSPVRQSPSPVRHVRAPTPSPVRSV-----SPAGRI 314
Qy 452 VVFCRVKGAESP--KVE-----WYREG--TLIEDSPDFRI----- 483
Db 315 STSPIRVKSPSPIRKAQVVTPGAELVLPWPRQEGYSATAEAAQMKETRVSTSAITEIRTEER 374
Qy 484 -----LQ-----KKPRMAEPPEEICTLVIA-----EV 505
Db 375 WEGRYGLQEQVTTISGAAGEVAAGAKEVRKEPKTPVPTVIIATDKAKEQERISTAREEI 434
Qy 506 FAEDSGCFTCTASKNG-----TVSSIAQLHVRGN----- 535
Db 435 SARHEQVHVSHQEQAEKRAEAVATVVAADQARVRSPEWTEQVDETYVVKKTKLEYGYKE 494
Qy 536 ---EDLSNNGSLHSANSTTNLAAIEPQPSPPHSEPPSV-----EQPPKPKLEGVL-VNH 585
Db 495 HAVKDEAQAQAEHHVATKEVKTVVVPEKHIPAAEKKEVHVSTEIKRETEAKIEKTIHIEH 554
Qy 586 NEPRSSSRIGLRVHFN-----LPEDDKGSASSEAGVVTTRQ-----TRPDSX 628
Db 555 PRPETAS-----PHFTVSKIAVAPKPDHTEVSIAGSAMATLEKELSATSAAQKITKPKVP 609
Qy 629 QERENGQATKTP-----PSFPVKKEPP-----PVLAKPKLDSTOLOLHNQVLLQHQLO 678
Db 610 POLKPEHVKIKPESAPQPFTEAAETIYKAHYDVETKEVDVS-----IKGEAVREDHLLL 665
Qy 679 NPPSS-----PKEFPFXMTVL-----NSNAPPVATTSX 707
Db 666 RKESEAKVTETARVPVPAETPVTPTLVWGLKNKTVTGESVTLCHISGHQPTVTWYR 725
Qy 708 KOVKAPSSQFSL-----ARPKYFPPTNTTAAATVAPSSSPVFTLSSTPQ 752
Db 726 EDYKIESMDQFQITFKAGLARLVIREAFEDSGRFTCTATNKAGSVSTCHLVKVKSEET 785
Qy 753 TIQRTVSKESLLVSHPSVQTK-----SPGGLSIOQLPPL-----GTEP 792
Db 786 ETRETISEKVVTSEKSVETKDVVMDVSAAAEEVSEPVPPFPIRPVHVHKLIEGGS-- 843
Qy 793 TPPPFTSIFSGNQFQPRCV---SPIVPSPTSR----- 822
Db 844 ---IIFECQVGGNPKPHVLMWKGKGVPLTTGYRYKSVKRYKRETKETGCKLEISMTFADDA 899
Qy 823 ---IQNPVAFSLSSVLPSPALPPIPNAMKLPDSAPSMPSOGLAKKNTKSPQ-----PV-- 871
Db 900 TTVIRNKFGEASA---TVSULEEADYEAYIKSQQEMMYQTQVTAIYQVEPKVAEVAPPISY 956
Qy 872 -NDNINIRETKNAVIR-DLGRKKITFSDVRPNQOEYKISSFEORLMEIEFRLERTPVD--- 926
Db 957 GDFDKVEYKEQALIRKKMAKDTVMVRTFVEDEEPIHSSFEERLIKEIELRIKTTIDELL 1016
Qy 927 ESDDIEOHDEIPTGCKIAPIDFKRLKHFRTVEGSPVFTCKIVGIPVYKVFYKDGKQIS 986
Db 1017 EEDGEEMMIDISEAIGAGFDRLKLNKRYTFEGTGVTFHCKTGYPLKPIAWYKDGKRI- 1075
Qy 987 KRNECHKRRREGDGTCSLHTESTSDDGNYTTMAANPOGRISCSGHLMVQSL----- 1039
Db 1076 RHGERYHMEVLQDGSASLRPLVLPDEGITYTFASNMKNKAICSAKLYPEVAPTATPG 1135
Qy 1040 -----PIRSRLTSAGQSHRGRSVORKEKPELOERFRFRPHFLOAPGD 1081
Db 1136 YMPCPEVMRRYRSTSPRSRSPARSPSPAR-RLDETDEGOL-ERLKPVEVLKPTS 1193
Qy 1082 MVAHEGLRCLDCVKVGLPPELTWLLNGQVPLPDASHKMLVRETGVHSLIDPLTORDA 1141
Db 1194 VKCSQGTARFDLVKVGGRMPETVYFHNGQVQVNDYTHKVIKEDGQSLLIIVPAMPEDS 1253
Qy 1142 GTVKCIATNKTKGNSFSLSVVAKEVKKAPVILEKLQNGVPEGHVPECRVIGMPPP 1201
Db 1254 GENAVIAQNAGRAGASVSTLSVEAKEDLVPRPREVERURNVSVREGSLRHMAVKATGNPNP 1313
```


Db	692	KERISAAKVAEPPQRPASEPHVVPVAKAPRV	IQAPSETHIKTTDDQGMHISQIKKTTDL	751
Qy	579	--EGVLVNHNEPRSSRIGLRVHF----	NLPEDDKGSEASSEAGVWTRQ-----	622
Db	752	TTERLVHDKRPRTAS-----PHFTVSKISV	PKTEHGEYASITAGSAIATLQKLSATSSA	806
Qy	623	-----TPDSSQERENGQAFTKPEPSFP	PKPEPPVVLAKP-----KLDSS	660
Db	807	QKITKSVKAPTVPKSETRVR---	ABEPTLPQFPPEADTPOTYKSEAGVEVK	862
Qy	661	TQLQOLHNQVL---LEQHQLONPPSP	PKPEFPMVILNSNAPPVTTSSXQVKAP	718
Db	863	TTVREERFEVLHGREAKVTETARVAP	VEIPV-----TPPTLVSLGKNVTIEGES	914
Qy	719	SLAPKVFPPSTNT	-----	733
Db	915	TLECHISGYDPSPIVTVWYREDYQIES	SIDPQITFQSGIARLMIREAFADSCRFT	974
Qy	734	AATVAPSS-----SPVFTLSSTPOT	IQRTVSKESLLVSHPSVQTKSPGGLSI	787
Db	975	AGTVSTSCYLAQVSEFEKETAVTEKFT	TEEKRFVESRDVWMTDT---SLTEE--	1028
Qy	788	GPTEPTPPP-----TFSPISGNQ	QOPRCV---SPIPVSPTSR-----	822
Db	1029	GPGEAAPYFITPKVVYOKLVEGSV	VFCQVGGNPKPHVYKWKSGVPLTT	1088
Qy	823	-----IONPVAFLSSVLP	SLPAIPPTNAMKLPRSAFSPMSQ	859
Db	1089	QTGECKLVISMTFADDAAGEYTV	IVRNKHGTSA---SASLLEEADYELLMK	1144
Qy	860	LAKKNTKSPQPVNDDNIRET-----	KNNAVIR-DLGKIKTTFS	903
Db	1145	-----TQVTAFOQEPEVGETAP	GFVYSEYEKEQALIRKKMAKDTVVV	1199
Qy	904	KISSFORLMNETEFLERTP-----	VDESDDIQHDEIPTCKCIAPIDPKL	959
Db	1200	HISSEERLIKETEYRIIKTTLEEL	EEDGEEKMAVDISESEAVESGFDL	1259
Qy	960	SPVTECKIVGIPVPKVPYFKDG	KQIKSKNEHCKMRREGDGTCSLH	1019
Db	1260	MGVTFCKKSGYPLPKIANYKDG	KRI-KHGERYQWDFLODGRASLRIP	1318
Qy	1020	MAANPOGRISCSGHLMVQ-----	SL-----PIR-----	1042
Db	1319	FASNIKGNATICSGLKVLVEPAA	PLGAPTYIPTLEPVYSIRSLSPRSVS	1378
Qy	1043	SRUTSAGQSHRGS---RVQERDK	PELPQRRFPHFLOAGDGMVAHEG	1099
Db	1379	ARMSPARMSPARMSPGRLEET	DESOL-ERLYKPFVFLKPSFKOLEG	1437
Qy	1100	PPPELTWLLNGQVLPDASHKML	VRETGVHSLLDPLTORAGTVKCIAT	1159
Db	1438	PMETTFWHDGQOIVNDYTHKKV	IKVEDGTQSLIIVPATPSDGEWTV	1497
Qy	1160	ELSWAKEVKAPVILEKLQNG	CVPEGHFVRLECRVIGMPPVFVW	1218
Db	1498	ILTVAEVEHQVKPMFEKLNK	YNIKEGSRLEMKVRATGNPNPD	1557
Qy	1219	ISWHQDITGYACLLIOPAKK	SAGWYTTLSAKNEAGIVSCTARL	1278
Db	1558	IRI-EGTKGEAALKIDSTVS	QSDSAWYTATINKAGRDTTRCK	1616
Qy	1279	PSGS 1282		
Db	1617	PRGT 1620		
RESULT 4				
T29757				
protein UNC-89 - Caenorhabditis elegans				
C:Species: Caenorhabditis elegans				
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999				
C:Accession: T29757				

Db 3470 ARNEAGEAL-----TTANFGIIRDSIP-----PEFTQKRLPVEQEOTLIDLKVTVIGT 3518
QY 972 PVPVYVFKDGKQKISKRNECKRRREGDGCPSLHIESITSDDDGNVTIMAANPQGRISCS 1031
Db 3519 PVPVNEVFKDKKPINIDNSHIFAKDESGHHTLTIKQARGEDGVVYTCRATNAGEAKT 3578
QY 1032 GHLMVOSLPPIRSRLTSAGQSHRGRSRVQERDKPELQERFFRPHFLQAPGDMVAHEGRCLR 1091
Db 3579 ANNAV-----QEEIEAPLFVQGLKPYEVEQKPAE 3608
QY 1092 LDCVSGSLPPELTWLLNGOPVLDPDASHKMLVR-ETGVHSELLDPLTQRDAGTYKCIATN 1150
Db 3609 LVVRVEGKPEVFKPDGVPPIAIDNOHVIEKKGENSGHTLVIKDNTNADFGKYTQATN 3668
QY 1151 KTGONSFSLSLV-----VAKEVKKAPVILEKLQNCVPGHGVRLRCRVLGMPPPV 1202
Db 3669 KAKDETGVGELKPKYKSEFEKQTAEEVK--PLFIEPLKETFAVGDTVVLECKVKNKSGHPQ 3726
QY 1203 FYWKKNDETTPCTRERISMHDQ---TTGYACLLIOPAKKSDAGWYTLTSAKNEAGIVSCT 1258
Db 3727 IKFFKNOPV-----ELGQHMOLVLEDGNIKLTIONAKKEDVGAYRCEAVNAGKANTN 3781
QY 1259 ARLDI 1263
Db 3782 ADLKI 3786

RESULT 5
S68235
myosin-light-chain kinase (EC 2.7.1.117), 210K, nonmuscle - chicken
N:Contains: myosin-light-chain kinase, 108K, smooth muscle; chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S68235; S67099; B44389; A44389; S28227; S78216; A35093; A25810; S11652
R:Watterson, D.M.; Collinge, M.; Lukas, T.J.; Van Eldik, L.J.; Birukov, K.G.; Stepanova,
FEBS Lett. 373, 217-220, 1995
A:Title: Multiple gene products are produced from a novel protein kinase transcription
A:Reference number: S68235; MUID:96033976; PMID:7589469
A:Accession: S68235
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1906 <WAT>
A:Cross-references: EMBL:X52876; NID:g992992; PIDN:CAA37056.1; PID:g992993
R:Shoenmaker, M.O.; Lau, W.; Shattuck, R.L.; Kwiatkowski, A.P.; Matrisian, P.E.; Guerra-S
J. Cell Biol. 111, 1107-1125, 1990
A:Title: Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides to
activity.
A:Reference number: A37099; MUID:90361738; PMID:2202734
A:Accession: A37099
A:Molecule type: mRNA
A:Residues: 649-1906 <SHO>
A:Cross-references: EMBL:X52876
R:Collinge, M.; Matrisian, P.E.; Zimmer, W.E.; Shattuck, R.L.; Lukas, T.J.; Van Eldik, L
Mol. Cell. Biol. 12, 2359-2371, 1992
A:Title: Structure and expression of a calcium-binding protein gene contained within a c
A:Reference number: A44389; MUID:92236611; PMID:1373815
A:Accession: B44389
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1695-1906 <COL>
A:Cross-references: GB:M88284; NID:g212237; PIDN:AAB53767.1; PID:g212238
A:Accession: A44389
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1750-1906
A:Cross-references: GB:M88283; NID:g211371; PIDN:AAA48647.1; PID:g211372
R:Yoshikai, S.I.; Ikebe, M.
Arch. Biochem. Biophys. 299, 242-247, 1992
A:Title: Molecular cloning of the chicken gizzard telokin gene and cDNA.
A:Reference number: S28227; MUID:93073972; PMID:1444462
A:Accession: S28227
A:Molecule type: mRNA

A:Residues: 1750-1906 <YOS>
A:Cross-references: EMBL:M96655; NID:g212744; PIDN:AAA49083.1; PID:g212745
A:Accession: S78216
A:Molecule type: DNA
A:Residues: 1750-1906 <YOW>
A:Cross-references: EMBL:M96987
R:Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurwitz, M.Y.; Kemp, B.E.; Means, A.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990
A:Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinas
A:Reference number: A35093; MUID:90192792; PMID:2315320
A:Accession: A35093
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 935-1438, 'O', 1440-1906 <OLS>
A:Cross-references: GB:M31048; NID:g212660; PIDN:AAA49069.1; PID:g212661
R:Guerriero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.
Biochemistry 25, 8372-8381, 1986
A:Title: Domain organization of chicken gizzard myosin light chain kinase deduced fro
A:Reference number: A25810; MUID:87157587; PMID:3030394
A:Accession: A25810
A:Molecule type: mRNA
A:Residues: 1258-1438, 'Q', 1440-1906 <GUE>
C:Genetics:
A:Introns: 1735/3; 1779/1; 1819/1
C:Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homo
C:Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phosphot
F:542-599/Domain: immunoglobulin homology <IMM1>
F:935-1906/Product: myosin-light-chain kinase, 108K, smooth muscle (from 5.5kb transcr
F:1098-1158/Domain: immunoglobulin homology <IMM2>
F:1451-1708/Domain: immunoglobulin homology <IMM3>
F:1459-1467/Region: protein kinase ATP-binding motif
F:1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transkript) #s
F:1808-1869/Domain: immunoglobulin homology <IMM3>

Query Match 8.1%; Score 556.5; DB 1; Length 1906;
Best Local Similarity 21.9%; Mismat. No. 2.9e-19;
Matches 237; Conservative 151; Prid. Matches 438; Indels 255; Gaps 36;

QY 267 QPREFTKLRSEVPCTRVQLDGVVGIPIPPQVVRWYCEGKELNSPDHIVQAG--NLH 324
Db 26 EAPFTLPPNRIRVOLGATARFEKGVRGYPEPQITVTRNGHPLPEG-DHYVVDHSIRGIF 84
QY 325 SLTIAAEFEEDTGRYSCFASNIYGTSDTSABEYIEGYSSSDSEGDPNKEENRIQKPNV 384
Db 85 SLVTKGVQEGDSGKYTCEAANDGGVRQVTVELTVEG-----NSLKYSLP 129
QY 385 SSPTTSAVTPPAVQAHLVAOPRVATIOQCQSPTNYLOGLDGKPII---AAPVFTKML 441
Db 130 SSATPGGRL--SVPPVEH-----RPSIWGESPPKFATKP 162
QY 442 ONLSASEGOLVVEFCRVKGAPSKVWYREGTLIEDSPFRILOKKPRSMAPPEICTLV 501
Db 163 NRVVREGOTGRFSCKITGRPQPOVWTK-----GDHLOQNERFNMFKTGIOYLE 214
QY 502 IAEVFAEDSGCFTASNKYGTVSSIAQLHVRGNEDLSNGLSHANSTTNLAAIEPQS 561
Db 215 IQNVQLADAGIYCTTVNSAGKASVSAELTVQG-----PDKT 251
QY 562 PPHSEPPSVEQPKP-KLEGVLVNHNEPRSSSRIGLRVHF-----NLPEDDKGSEASSEA 615
Db 252 DTHAQP--LCMPKPPTTLATKAIENSDFKQATNSGIAKELKSTSTELMWETKDRLSAKKE 309
QY 616 GVVTRGTRDPSXQERENGQATKTPPEPSFPVKEPPVLAHPKLDSTOLOQLHNVLEQH 675
Db 310 TFFYTSREAK-DGKQGG-QNEANAVPQESRGTKGPQV-----LQKTSSTITLQAV 357
QY 676 QLQNPFPSPKPEFPFXMTVLNLSNAPPVATVTSXKQVKAPSSQTFSLARPKYFFPSTNTTAA 735
Db 358 KAQPEPKAEFQ-----TTFIROAE-----DRKRTVQPLMTTIT- 390
QY 736 TVAPSSSPVFTLSSTPQTQITQVTSKESILLVSHPSVQTKSPGL--SIQNEPLPFGPTEPT 793
Db 391 ----QENPSLTGQVSPR-----SRETENRAGVRKSVKEEREP----- 424

Db 1276 KIAASLPYFIELKPKINNVVEGATLSIQADLNGSPI-----PEVV--WLKDNSELVES 1327
Qy 768 PSVOTKSPG-----GLSTQNE----- 783
Db 1328 DRIOHKDGVNQLLRVDVGLDESGTYTITAEENKGIQRTVEVSVTKSKEVKEKKKK 1387
Qy 784 -----PLPPGTEPTPTPPFTSIPS-----GNQFQPRCVSP---IPVSP 819
Db 1388 VEKKDEGKKKPGRLPPSCASKTEQVTMAFDAPSEGPAISYEVERCPCDQREWVSGS 1447
Qy 820 TSRQNPVAFUSSVLPSLPAIPPTNMXLPR-----SAPSMPSOGCLKAKNTKSPQ----- 870
Db 1448 TKSLELEIKGLTPNTEYIFRVAGNKKOGLGEWSEMTSLTKTASVGAQAPQFTISPSQKIIA 1507
Qy 871 -----VNDDNI-----RETNAV 883
Db 1508 NRDEFEIATVEFGTPTPSVKWYKXENLOIVDEKIDVATTSTSSILNLKSOEENGTFNCL 1567
Qy 884 I-RDLGK-----KITFSDVRNQOEYKISSPEORLM-----NE-----I 916
Db 1568 IENELGQASASCQVTIFNKPASLOSTPDHSLERNLPTLQKALNESAQAQOQIMLTCTRI 1627
Qy 917 EFRLEPTPVDESDE-----IQHDEIPTGKC----- 942
Db 1628 SSRSESTVAMFKDDERIESAGRYELSSDKKSNHKLVLCHAVQSDTGYKRYCVVTKNGYAE 1687
Qy 943 -----TAPIDFKRLKHFRTVEGSPVTFCKIVGPVYKVVWPKDQKQISKRN 989
Db 1688 SECNVAVEDVTKFTAPSFATLSSTAILGNHITLECKVEGSPAPEVSWTKDGERIS--TT 1746
Qy 990 EHKMRREGDGTCSLIHIESTTSDDDGNYTIIAANPQGRISCSGHLWQSLPIRSRLTSAG 1049
Db 1747 RRIQTQDENGCKLSISKAESDDMGVYVCSATSVAGVDSTSSVMIAK-----TTGT 1799
Qy 1050 QSHGRSRVQRDRKEPQERFRPHFLQAPGMV-AHEGRICRLDCKVSGVLPPLPELWLL 1108
Db 1800 DSHLVIAQTAD-----EKHEPRFTRAPPSLIEVNESGQFTLIKAVGPKPTVTWLK 1852
Qy 1109 NGQPVLPDASHKMLVRETGVHSLIIDLPTORDAGTYKCIATNKTGQNSFSLSVVAKE 1167
Db 1853 DGREILTRNRIHFVTDGSGHSLIAECVSVKTSIGFISCKAENPNG--TVIAETQVIVQR 1910
Qy 1168 VKK-----APVILEKLQNGVPEGHVLECRVIGNPPVFFVYKKDKNETIPCTRERIS 1220
Db 1911 MKPANLANVAPKFTIPLTDGIVNGHPTLTSCNVNVTGSPTELEW----IYDDSGHKIN 1966
Qy 1221 MHQDTTGY-----ACLLIQPAKSDAGWYTLAKNEAGIVSCTARLDIYAQWHHOIP 1272
Db 1967 LTSSTDTWTECRFGKVAELKSERVLREQRTYQCIATNSSGQATTCQCYLLVGLSDEPAG 2026
Qy 1273 PP-----MSVRPSGSRYGSLT-----SKGLDIFSASFSSMESTMV 1306
Db 2027 PPRFVKCLQDTWTPLKESIEFSVELAGPTPDLTWYHNEKKINEGKDVKITFPS-DTITSV 2085
Qy 1307 YSCSSRSV 1314
Db 2086 LSIKNVSL 2093

RESULT 9

T29340

hypothetical protein F21C10.7 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29340

R:Du, Z.; Gattung, S.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of *C. elegans* cosmid F21C10.

A:Reference number: 220610

A:Accession: T29340

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2541 <DUZ>

A:Cross-references: EMBL:U5364; PIDN:AAA97973.1; GSPDB:GN00023; CESP:F21C10.7
A:Experimental source: strain Bristol N2; clone F21C10
C:Genetics:
A:Gene: CESP:F21C10.7
A:Map position: 5
A:Introns: 50/3; 91/3; 153/3; 184/3; 215/1; 254/3; 350/1; 392/3; 538/1; 589/3; 625/3;

Query Match 7.38: Score 502; DB 2: Length 2541;

Best Local Similarity 18.18; Pred. No. 1.9e-16;

Matches 320; Conservative 206; Mismatches 536; Indels 706; Gaps 60;

Qy 59 DLPDLSAFLSQEELDESYNLARLAINY-----DPLEKA----- 91
Db 905 DSTPMKTVOESQELQEKVKLVKEAIEVQCPKIEDVAVAKARDIAVKTOISRAVEKQVIRE 964
Qy 92 -DETQAKRLSPDMQKISPNLSPFNFCQDNPRSPSTSKESQPAKRPQYCSQTQSKVVF 150
Db 965 MAETLERKAISATFELSKPQOTIE-----ETKTIELTSLEQTITLTPEEKQELIILKKI 1019
Qy 151 LKNAADTFEELSSFLFKSHSSKRIRPRACKNHKSKLESQNKVMQ-----ENSSSFSD 201
Db 1020 VEEIQMQEETIEIILQVDK---TPKTVQSEETKIKKVEIQOIVVAVQTRLEDAERF-- 1074
Qy 202 LSERRERSVPIPIPADTRDNEVNAHQEQEAKRREARQAASEAAGGDTTPGSSPSLLY 261
Db 1075 -----RKDETFKTQETMTKQOQVQQLVKELHERAASQNIIVVRKQEE 1117
Qy 262 EEPGLQPPRTQKLRREVPEGTRVQLDCIVVGIPPOVVRWYCEGKELENSPDTHIVQAG 321
Db 1118 ERSVQAPQIITQIKDDVYDEGCRYEFESARINGEPEPTISWLKDGIDVKSNDYRQEVYN 1177
Qy 322 NLHSITTAEAPEEDTGRYSCFASNIYGDSTSAEYITEGVSSSDSEGDPNKEEMNRQKP 381
Db 1178 GVAILVTEESTIEDTAETVKAATNTGTGASSASLIVKRSR----- 1219
Qy 382 NEVSPPTTSVAVIPVPAQAOHLVAQPRVATIQCCQSPNTYLOGLDGKPIIAAPVFTKML 441
Db 1220 -----MSSAILEDKPR-----FVKQM 1236
Qy 442 QNLSASGQLVVFECRVKGPAPKPVVYREGTLIEDSPDFRILQKKPRMAPEPEICTLV 501
Db 1237 QSVQVNEGETARLDCVVGKPEPEITWFKETAVKESR--RV-----HLTFSGDHQOMI 1288
Qy 502 IAEVFAEDSGCFTCTASNKYGTSSIAQLHVRGNEDLSNNGSLHSANSTTNLAATPEOPS 561
Db 1289 IDKIVPLDTGIYTVRAKNVHGEVANFCOLRV-----VPKQK 1324
Qy 562 PPHSEPPS-----VSEOPP--KPKL-----EG-----VLVNHNEPRSSRIGLRVH--FN- 601
Db 1325 PPQTPPKRPTIQRPPVQIPALNTTWOEGETATLQVFSYGEKPK-----RVHWKFND 1378
Qy 602 -----LPEDDKGSE-----ASSEAG-----V 617
Db 1379 SPVOTSSQVQISEQEDGWSRLTIQOISPVNAGMTVVAENIEGEAVTGATVHVQPSLKR 1438
Qy 618 VTTRQTRPDSXQERFNGQATK-----PEPSFPVK----- 647
Db 1439 VTTEHLLQEDMNEQIGQVPVQQTIIITKKSQELETRKTDFFQPREPEKIVDEKRWVLEVEQ 1498
Qy 648 -----EPPPVVLAKPK-----LDSTQLQQLHNOVLLBQH 675
Db 1499 HFEEHLSQRSISPVQIREVRKTESQQRWIDTIDEINSPMREVDITETRTVSGMDQYSSH 1558
Qy 676 QLQNPSPSSPKPEFFXMTVLNSN-----APPV---TTSXKQVKAPSSQTFSLAR 722
Db 1559 NVHEP---SPRPVGYHTTTTTTINSHIGQSHEPVQPVNMGSRSTSSNETVKTNIATIRQSP 1615
Qy 723 PKYFFPSTNTTAAATVAPSSSPVFTLSSTP--QTIQRTVSKESLL-----VSHP 768
Db 1616 QREVEYARTP-----GPTSEHVATIRKTPVRETHQSSITTKAPSLNVAKIRQSPVPSHI 1671
Qy 769 SVOT-KSP-----GGLSIQNEPLPPGPTPTPTPTPSIP--- 802

Qy 918 ----FRLRTPVDESDDDEIQHDEIPTGKICIA-----PIFDKRLKHF 954
Db 2996 KSKLYIMQATPEDAD-----SYSCIAVNDAGGAEAVFQVTVNTPPKIFGDSFSTT 3045
Qy 955 RVTEGSPVTFCKIVGIPVVKVYFKDGKQISKRNECHKMRREG-----DGTCSLHIESTT 1010
Db 3046 EIVADTTLEIPCRTEGIPPPISWFLDGKPI-----LEMPGVYTKQGSLSLRIDNIK 3097
Qy 1011 SDDGNTYMAANPQGRISCSGHLWVQSLPIRSLTSAGSHRGRSRVQERDKPELQERF 1070
Db 3098 PNQGRYTCVAENKAGRAEQDTYVEI-----SEP----- 3126
Qy 1071 FRPHFLOAGDMVAHEGRCLRDCKVSGLPPELTWLLNGQVLPDPASHKMLVRETGVHS 1130
Db 3127 --PRVWASEVMRVVEGRTQIRCEVFNPEPVNWLKDGEPYTS-----LLQFSTKLSY 3180
Qy 1131 LLIDPLTORDAGTYKCIATNKGTGONSFSLSLVAKKAPVILEKLQNGCVPEGHVPR 1190
Db 3181 LHLRETTLADGGTYTCIATNKAGESQTTDV-----EVLVPPRIEDEERVVLQGEKNTYM 3235
Qy 1191 LECRVIGMPPPVFVYKKDNETIPTCTRERISMHQDTTGYACLLIOPAKKSDAGWYTLGSAKN 1250
Db 3236 VHCQVTRGPVYVTKRNGKEI-----EOPN-----PVLHIRNATRADEGKYSCIASN 3283
Qy 1251 EAGVSCITARDIYQA 1266
Db 3284 EAGTAVADFLIDVETK 3299
RESULT 11
T43290
hemictinin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43290; T20993; T24734
R:Vogel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemictinin is required for hemidesmosome mediated cell adhesion and germ-
A:Reference number: 222396
A:Accession: T43290
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <VOG>
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20993
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN000028; CESP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN000028; CESP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1
; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1
Query Match 6.8%; Score 468.5; DB 2; Length 5198;
Best Local Similarity 21.58; Pred. No. 1.9e-14;
Matches 248; Conservative 161; Mismatches 448; Indels 299; Gaps 48;

Qy 268 PPR--FTQKLRRESVECTRVQLDCIVVGIPIPPQVRWVCEGK--ELENSPDHIVQAGNL 323
Db 2286 PPTIIMLDKDNKTAVEHSTVTLSCPATGKPEPDITWFKDGEATHENIAD--IIPNGEL 2343
Qy 324 --HSLTTAAFEEDTGRYSCFASNIYGTDSYSAEIIYEGVSSSDSEDPNKEENNRITQKP 381
Db 2344 NGNOLKITRIKREGAGKYTCEDANSAGSVEQDVNVNVTIPKIEKDGIPSDYESQ--QNE 2401
Qy 382 NEVSSPPTTSVAVIPPA---VPOAHLVAOPRVATIOQCQS-----PT 420
Db 2402 RVVJSCPVYAR--PPAKITWLKAGKQLQSDKFKVTSANGQKLYLFKURETOSSKYTCTIAT 2459
Qy 421 NYLOGLDGK----PIIAAPVETK--MLQNLSASEGQLVVFEFCRVKGPAPSKPVKEVREGTL 474
Db 2460 NE-AGTKRDPRKFKVSMVLVAPSDPEPNIVRRITVNSGNSTLHCPAKGSPSPITWLKDGNA 2518
Qy 475 IEDSPDRFRILOKKRPSMAEPBEEICTLVIAEVAEDSGCTCTASNKYGTGVSSIAQLHVRG 534
Db 2519 IEPNDRYVFFDAGRQ-----LQISKTEGSDQGRYTCIATNSV-----G 2556
Qy 535 NEDLSN-----NGSLHSANSTTNLAALIEPQSPPHSEPPSVEQPPKPLEGLVLN 584
Db 2557 SDDLENTLEIIPVIDGERREA-----VAVIEGFSSELFCDNSSTGVVDVEMQDGLTIN 2611
Qy 585 HNEPRSSSRIGLRVHFNLPEDDKG-----SEASSEAGVVVTRQTRPDSAXQERFNGQATKTP 640
Db 2612 QDTLRGDSFI-----QIPSSGKKMSLSARKSDSGRYTCIVRNP-----AGEARKLF 2658
Qy 641 EPSFPVKPEPPVLAOKPLDSTQLQOL-----HNOVLLLEHQHQLQNPPSS 684
Db 2659 D--FAVNDPPSI--SDELSANIQTIVPYVPEINCVVSGSPHPKV---YVLFDDKPLE 2710
Qy 685 PKERFFXMT-----VLNSNAPPVAVT--TSXKQVKAQSSQTFESLAR---PKYFPSTNTT 733
Db 2711 PDSAAIYELTNNGETLKVRSQVEHAGTYTCEAQNNGKARKDFLVRVTAPOHFEKEREEV 2770
Qy 734 AATV-----APSSSPVFTL--SSPTQTIQRTV-----SKESLLVSHSPSVQTKSP 775
Db 2771 VARVGDMILLTCNAESSVPLSSVYVHWAHDESVQNGVITSKYAANEKTLNVT--NIQLDDE 2828
Qy 776 G-----GLS-----IQNEPLPP-----GTEPTEPPPTFFSI 801
Db 2829 GFYCTAVNEAGITKFKFLIVETPYFLDQOKLYPIILGKRLTLDGCSATGTPPTTILFM 2888
Qy 802 PSGNQFOPRCVSPIPVSPTSRIQNP---VAFLLSVLSPLPAIPPTNAMXLPRSPAPSPSQ 858
Db 2889 KDGRLNESDEVDI--IGSTLVIDNPQKEVEGRTYCIENKAGRSEKDMVAVLLPKLSK 2947
Qy 859 GLAKKNTKSPQPVNDNIRETKNAVIRDL-GKKITFSVDRPNQOEYKISSFEQRLMNEIE 917
Db 2948 EWINVEVQAGDPL-----TLECPIEDTSGVHITWS-----RQFGDKGQOLDMRAQSSSD 2995
Qy 918 ----FRLRTPVDESDDDEIQHDEIPTGKICIA-----PIFDKRLKHF 954
Db 2996 KSKLYIMQATPEDAD-----SYSCIAVNDAGGAEAVFQVTVNTPPKIFGDSFSTT 3045
Qy 955 RVTEGSPVTFCKIVGIPVVKVYFKDGKQISKRNECHKMRREG-----DGTCSLHIESTT 1010
Db 3046 EIVADTTLEIPCRTEGIPPPISWFLDGKPI-----LEMPGVYTKQGSLSLRIDNIK 3097
Qy 1011 SDDGNTYMAANPQGRISCSGHLWVQSLPIRSLTSAGSHRGRSRVQERDKPELQERF 1070
Db 3098 PNQGRYTCVAENKAGRAEQDTYVEI-----SEP----- 3126
Qy 1071 FRPHFLOAGDMVAHEGRCLRDCKVSGLPPELTWLLNGQVLPDPASHKMLVRETGVHS 1130
Db 3127 --PRVWASEVMRVVEGRTQIRCEVFNPEPVNWLKDGEPYTS-----LLQFSTKLSY 3180
Qy 1131 LLIDPLTORDAGTYKCIATNKGTGONSFSLSLVAKKAPVILEKLQNGCVPEGHVPR 1190
Db 3181 LHLRETTLADGGTYTCIATNKAGESQTTDV-----EVLVPPRIEDEERVVLQGEKNTYM 3235
Qy 1191 LECRVIGMPPPVFVYKKDNETIPTCTRERISMHQDTTGYACLLIOPAKKSDAGWYTLGSAKN 1250

A:Experimental source: var. Bristol
R:White, S.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20442
A:Accession: T27934
A:Status: preliminary; translated from GE/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 'MGIPGKCKQ', 19-6839 <WIL>
A:Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a
A:Experimental source: clone ZK617
R:Harris, B.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20458
A:Accession: T28030
A:Status: preliminary; translated from GE/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 'MGIPGKCKQ', 19-6839 <W12>
A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a
A:Experimental source: clone ZK829
C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C:Genetics:
A:Gene: unc-22; CESP:ZK617.1a
A:Map position: 4
A:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 582/3; 669/3
152/3; 669/3; 6776/1; 6808/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase
F:806-898, 899-990, 991-1083, 1084-1175, 1178-1273, 1474-1567, 1770-1864, 2066-2158, 2358-2450, 2
96-570, 6263-6386, 6386-6478, 6541-6649, 6742, 6745-6838/Region: motif 2
F:1274-1372, 1373-1473, 1568-1670, 1671-1769, 1865-1964, 1965-2065, 2159-2258, 2259-2357, 2451-2
23, 4215-4313, 4314-4415, 4416-4516, 4612-4710, 4711-4811, 4908-5009, 5010-5109, 5110-5210, 5399-
F:5940-6197/Domain: protein kinase homology <KIN>
F:5948-5956/Region: protein kinase ATP-binding motif
F:5971/Active site: Lys #status Predicted

Query Match	5.8%;	Score 402;	DB 2;	Length 6839;
Best Local Similarity	18.3%;	Pred. No. 4,7e-11;		
Matches	288;	Conservative 153;	Mismatches 532;	Indels 600; Gaps 50;
Qy	204	ERRSSVPIPADTRDNEVNHAEQ-QEAKRREAQAASEAAG-GDTTGCSSPSSLIYY	261	
Db	5149	EIREYGSPLTWASDYNVREPFVTKLREDFNEFFRVAINAAGKGLPSLPSPKIK--	5206	
Qy	262	EELGQPPRFQTKLRSREVPCTGRVQLDCIVVGVIPPQPVQVYCGEKLLENSPDTHIVQAG	321	
Db	5207	QESGGRSQIIVVKEDPTAQPYNRRAVFTCEAVGRPEPTARWLNRGRELPESSRYRFEASD	5266	
Qy	322	NLHSLTIAAEFEEDTGRYSCFASNIYGDTSAS-----	354	
Db	5267	GVYKFTKEVWDIDAGEYTVESNPYSGDSTATANLVQAPPVIEKDVPTNLTLPSGDLVRL	5326	
Qy	355	EIYIEGVS-----	366	
Db	5327	KIYFSGTAPFRHSLVNLNREIDMDHPTIRIVEFDHILITIPALSVREAGRYEYTVSND	5386	
Qy	367	-----EGDP-----	370	
Db	5387	GEATTGFWLNTGLPEAPQGPLHISNIGPSTATLSWRPPVTGGSKITSYVVEKRDLSKD	5446	
Qy	371	-----NKEBMNRI-----	400	
Db	5447	EWVTVTSNVKDMNYIVTGLFNHEYEFRRVSAQNENGICAPLVSEHPIIARLPFPDPTSPL	5506	
Qy	401	AQHLV-----	422	
Db	5507	NLEIVGVGGDYVTLVSWQRPLSDGGGRLRGYIVEXQEEHDEHFRCNQNPSPNNVNVNPL	5566	
Qy	423	-----LQGLDGKPIIAA-----	457	
Db	5567	IDGRKYRYRFAVNDAGLSDLAEILDQTLFQASGSGEGPKIYSPLSDLNVEVGRCVTFECE	5626	
Qy	458	VKGAPSPKVEWYRGTLTIEDSPDFRILQKPRSMANPEICTLVIAEVFAEDSGCFTCTA	517	

Db	5627	ISGSPREYRWFGCKELVDTSKYTLINKGQKV-----LIINDLTSDADBEYCR	5677
Qy	518	SNKYGTVSSTAQHLVRGNEDLSNNGSLHSANSTTNLAIE---POPSPHPSEPPSVQPP	574
Db	5679	TNSGTRSTRANLRITKPRVFIPPKYHGYEAKGETIELKIPKAYPOGEARWK---	5735
Qy	575	KPLEGVLVNHNPRSSRIGLRVHNLPEDDK-----GSPASSEAGVWTRTRP	625
Db	5736	---DGEKIENN-----KFSITDDKFATLRISNASREDYGEYRVVWVENS	5780
Qy	626	DSXQERENGATKPE-PSFPVKE-----PPPVL-----	653
Db	5781	DSG--TVNVTVADVPFPRPIIENILDEAVILSWKPPALDGGSLVWYTIETKREAMGS	5838
Qy	654	---AKPKLSDTOLQ-----QLHNQVLEHOHLQNP--PSSPKFEP-----	689
Db	5839	WSPCAKSRYYTYTIEGLRACKQYEFRIIAENKHGQSKCEPTAPVLIPGDERKRRGYD	5898
Qy	690	---FXMTVLNSNAPPAVT-----	704
Db	5899	DEQKIVRGKGTSSNDYVDFWIKQYYQPQVEIKHDHVLVDHYDHIHELGTAGFVWH	5958
Qy	705	---TSXQVKVAPSSQTFSLARP-----YFF	727
Db	5959	VTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHADFEDDNEMWMI	6018
Qy	728	PSNTTAATVAPSSPV-----FTLSSTQTIQTV--SKE	761
Db	6019	MSGGELFEKVADEHNKSEDEAVEYMRQYKGLCHMHENNYVHLDLKPENIMFTTKS	6078
Qy	762	SLLVSHPSVOTKSPGGLSQNEP---LPPGPTPTPPPTFTSPG---NQOPRCVS	813
Db	6079	LKLIDE-----GLTAHLDPKQSVKTTGTGTAFAAEVAGKPVGYTDMNSVG	6129
Qy	814	PIPVSPTSRI--QNPVAFLSV-----LPSLPALPPTNAMXL-	848
Db	6130	YILLGSLSPFGGENDDETLRNKSCDMNDDSAFGSISGDKDFIRKLLADPNTRM	6189
Qy	849	-----PRSAFSPSQGLAKNKTSPQVNDNIRETKA---VIRDLGKKTFS	896
Db	6190	QALEHPWLITGNAPRDSQIPTSRYTKI-----RUSIKTKYDAWPEPLPP	6244
Qy	897	RPNO-QEYKI--SSFEORLMEIEFRLETPVDESDEIQHDEIPTGKCIAPIDFK	954
Db	6245	RKHPRQEYSIRDAFWDRSEAOPRFIVK-----PYGT-----	6275
Qy	955	RVTEGSPVTTCITVGPVPKYVFWKDGKQISKRNEHCKMRREGDGTCSLH	1014
Db	6276	EVGEGSANSYCYRIASSPPVVTWMDRELKQSVKY--MKRYNGNDYGLTIN	6333
Qy	1015	GNVTIMANPQGR-----ISCSGHLMWQSLPIRSRLTSAGOSHRSRVOER	1068
Db	6334	GEYTVRAKNSYGTKEEIVFLNVRH-----SEPLKFEPELPMKAKASPR	6384
Qy	1069	RFFRPHFLOAPGDMVAHEGRCLRCDCKVSGLPPPELTWLLNGOQVLPDASH	1128
Db	6385	RRSAPPTFHLNRLIQNHOCKLTCLOGNPNTIEMWKDGHV--DEDRQVVSFR	6442
Qy	1129	HSLLIDPLTORDAGTYKCIATNKTGNSFSELSV-----	1163
Db	6443	CSLEIFNARVDDAGTYVTATNDLGDVSECVLTVQTGGEPIPRVSSFRPR	6502
Qy	1164	-----VAKEYKKA-----PVILEKLQNCVPEGHVPR	1190
Db	6503	GTDVERSHVSADMRRRSILRDVSPDVRSAADDLTKITNELSFQAQSDSE	6562
Qy	1191	LECRVIGMPFVYWKXDNETIPTCTRERISMHDOTTGYACLLIQAPKSDAG	1250
Db	6563	FSAVSGQPELIEWLHNGERISESDSRFRASY-VAGKATLRISDAKKSDE	6621
Qy	1251	EAGTVSCTARLDI	1263
Db	6622	SAGOEQTRATILV	6634

RESULT 15

T27935 hypothetical protein ZK617.1b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000

C:Accession: T27935; T28031

R:White, S.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20442

A:Accession: T27935

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7160 <N1>

A:Cross-references: EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b

A:Experimental source: clone ZK617

R:Harris, B.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20458

A:Accession: T28031

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7160 <N1>

A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b

A:Experimental source: clone ZK829

C:Genetics:

A:Gene: CESP:ZK617.1b

A:Map position: 4

A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 59

3067/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3

C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 5.8%; Score 402; DB 2; Length 7160;

Best Local Similarity 18.3%; Pred. No. 5e-11;

Matches 288; Conservative 153; Mismatches 532; Indels 600; Gaps 50;

QY 204 ERRSRSSVPIPADTRDNEVNHAEQ-GEAKRREAEQAASEAAG-GDTTPGSGPSSLIYY 261

Db 5470 EIREVGSFLWVADSYNNREPEFTVDKLRNFEDYFRVAINAAGKIGPSLPSPGIKI-- 5527

QY 262 EEPGQPPRFQTKLRSRVPGTGVQVDCVIGPPQVQVRYGCKELENDSPDHIHQVAG 321

Db 5528 QESGSRQIVVVKPDEDTAQPNRRRAVFTCEAVGRPEPTARLNRGLRELPESSRYFEASD 5587

QY 322 NLHSITIAEAEEDTGRVSCFASNIYGDSTSA----- 354

Db 5588 GYVETKEWIDAGETVVEVSNPYGSDTATANLVQVAPPVIEKDVPNTILPSGDLVRL 5647

QY 355 EYIEGVG-----SSDS 366

Db 5648 KIYFSGTAPFRHSLVLRNEEDMDHPTIRIVEFDHILITIPALSVREAGRYEYTVSND 5707

QY 367 -----EGDP----- 370

Db 5708 GEATGFWLNTGLPEAPQGLHISNIGPSPATLSWRPPVTDGGSKITSVYVEKRDLSKD 5767

QY 371 -----NKEEMNRI-----QKNEVSSPPTTSAVI-----PPAVPQ 400

Db 5768 EMTVTSNVKDMNVIVGLFENHEYEFVRSQAQNGENGICAPLVSEHPITARLPDPPTSP 5827

QY 401 AQHLV-----AQPRVATIO-----AQCS-----PTNY----- 422

Db 5828 NLEIVQGGDYVTLVSWORPLSDGGRLRGYIVEKEEHEDEWFCNQNPSPNNVNPVL 5887

QY 423 -----LQGLDGRPIIAA-----PVFTKMLONLSASGQLVVECR 457

Db 5888 IDGRKRYRVFAVNDAGLSDAIEDQTLFQASGSGEGEKIVSPSLDNEEYGRCVTFECE 5947

QY 458 VKGAPSPKVEWYRGTLIEDSPDFRILOKPKRMAPEICTLVIAEVAEDSGCFTCTA 517

Db 5948 IGSPPREYRWFCKCKELVDFTSKYTLINKGDKV-----LIINDLTSDDADEYTCRA 5999

QY 518 SNKYGTVSSIAQLHVRGNEDLSNGLSHSANSNTNLAIE---PQSPPHSEPPSVEQPP 574

Db 6000 TNSGSTRSTRANLRKTKPRVFIIPKYHGGYEAQKGETIELKIPVKAYPQCEARWTK--- 6056

QY 575 KPKLEGVLVNHNEPRSSSRIGRLRVHFNLPEDDK-----GSEASEAGVVTTRQTRP 625

Db 6057 ----DGEKIENS-----KFSITDDKFKATLRISNASREDYGEYRVVSENSVGS 6101

QY 626 DSXQERENGQATKPE-PSRPVKE-----PPVL----- 653

Db 6102 DSG--TVNVTVADVPPEPRPFIIENILDEAVILSWKPPALDGGSLVNYVTIEKREMGGS 6159

QY 654 ----AKPKLDSTOLQ-----QLHNQVILQHQLOQNP--PSSPKPEF----- 689

Db 6160 WSPCAKSRVYTTTIEGLRAGKQYEFRIIAENKHGSKPCEPTAPVLIPGDERKRRRGYDV 6219

QY 690 -----FXMTVLNSNAPPAVT----- 704

Db 6220 DEQGIKVRGKTSSNYVDYFQVVEIKKHVDHLDHYDIHEELGTGAFGVVHR 6279

QY 705 -----TSXKQVKAPSSOTSFLARPK-----YFF 727

Db 6280 VTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAFEDDNEWMVIYEF 6339

QY 728 PSTNTTAATVAPSSSPV-----FTLSSTPQTQRTV--SKE 761

Db 6340 MSGGELFEKVADEHNKXSEDEAVEYMRQVCKGLCHMHNNVYVHLDLKPENIMFTTKRSNE 6399

QY 762 SLLVSHPSVQTKSGGLSIQNEP-----LPPGPTPEPPPTFTFIPSG---NQQPRCVS 813

Db 6400 LKLDF-----GLTAHLDPQSKVKVTGTGTAFAAEPVAEGKPVGYTDMMSVGVL 6450

QY 814 PIPVSPTSRI--QNPVAFLLSV-----LPSLPAIPPTNMXKL- 848

Db 6451 YILLGLSPFGGENDDELNRVSKCDWNMDDSAFSGISGDKDFIRKLLADPNTRMTIH 6510

QY 849 -----PRSAPSMPSQGLAKNTKSPQVNDNDNIRETKNA---VTRDLGKKTITSDV 896

Db 6511 QALEHPWLTPGNAGRDSQIPSSRYTKI-----RDSIKTKYDAMPEPLPLGLRISNSSL 6565

QY 897 RPNQ-OEYKI--SSFEQRLMNEIEFRLERTPVDESDDDEIQHDEIPTGKCIAPIDKRLKHF 954

Db 6566 RKHRPQEYSIRDAFDWSEAQPRFIVK-----PYGT----- 6596

QY 955 RVTEGSPVTFCKIVGIPVKVYWFKQKQISKNEHCKMRREGDGTCSLHISTSDDD 1014

Db 6597 EVGQGSANFCRVIASPPVYVTHKDDRELKQSVKY--MKRYNGNDYGLTINRVKGGDK 6654

QY 1015 GNYTIMAANPOGR-----ISCSGHLMVQSLPIRSRLTSAGQSHRGRSRVQERDKEPLQE 1068

Db 6655 GEYTVRAKNSYGTREEIVFLNVRH-----SEPLKFEPLPMKAPSPRVEE-----FKE 6705

QY 1069 RFFRPHFLOAPGDMVAHEGRLCLRDCKVGLPPPELFTWLLNGQVLPDASHKMLVRETGV 1128

Db 6706 RRSAPFTFHLNRLLIOKNHOCKLTCLOGNPNTIEMKMDGHEV--DEDRVQVSRSGV 6763

QY 1129 HSLLDPLTQORDAGTYKCIATNKTGQNSFSLELV----- 1163

Db 6764 CSLEIFNARVDAGTYVYVATNDLGDVSECVLIVQTKGGEPIPRVSVSFRPRRAYDTLST 6823

QY 1164 -----VAKEVKA-----PVILEKLQNGCVPEGHVPR 1190

Db 6824 GTDVERSHYADMRRLSLRDVSPDVRSAADDLTKTKITNELPSTAQLSDSETEVGSGAE 6883

QY 1191 LECRVIGMPPVFWKKDNETIPTCTRERISMHQDTTGYACILLIQPAKKSAGWTLNLSKN 1250

Db 6884 FSAAVSQPEPLIEWHLNGERISESDSRFRASY-VAGKATLURISDAKKSDGQYLCSAN 6942

QY 1251 EAGIVSCTARLDI 1263

Db 6943 SAGOEQTRATLTV 6955

Search completed: November 30, 2002, 19:07:41
Job time : 79 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 18:51:30 ; Search time 18 Seconds
(without alignments)
3041.597 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence: 1 MQDDSIESTSISQLLRESY.....MESTWVYSCSRSVVESDEL 1320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	556.5	8.1	1906	1	KMLS_CHICK	P11799	gallus gall
2	546	7.9	1914	1	KMLS_HUMAN	Q15746	homo sapien
3	392	5.7	4393	1	PGBM_HUMAN	P98160	homo sapien
4	375.5	5.4	3707	1	PGBM_MOUSE	Q05793	mus musculus
5	364	5.3	1176	1	KMLS_BOVIN	Q28824	bos taurus
6	352.5	5.1	1147	1	KMLS_RABIT	P29294	oryctolagus
7	322	4.7	2481	1	UN52_CAEEL	Q06561	caenorhabditis
8	317	4.6	1465	1	MYM2_HUMAN	P54296	homo sapien
9	315	4.6	2012	1	DSCA_HUMAN	O60469	homo sapien
10	307	4.5	1447	1	DCC_MOUSE	P70211	mus musculus
11	301.5	4.4	1450	1	MPSF_CHICK	Q02173	gallus gall
12	300	4.4	1447	1	DCC_HUMAN	P43146	homo sapien
13	296.5	4.3	1709	1	SN_HUMAN	Q9B222	homo sapien
14	282.5	4.1	1666	1	MYM1_MOUSE	G62234	mus musculus
15	278	4.0	837	1	NCM2_MOUSE	O35136	mus musculus
16	272	3.9	725	1	NCA2_MOUSE	P13594	mus musculus
17	272	3.9	1115	1	NCA1_MOUSE	P13595	mus musculus
18	271	3.9	837	1	NCM2_HUMAN	O15394	homo sapien
19	271	3.9	853	1	NCA1_BOVIN	P31836	bos taurus
20	268.5	3.9	1070	1	PTK7_HUMAN	Q13308	homo sapien
21	267.5	3.9	1040	1	AXO1_HUMAN	Q02246	homo sapien
22	267	3.9	761	1	NCA2_HUMAN	P13592	homo sapien
23	264.5	3.8	848	1	NCA1_HUMAN	P13591	homo sapien
24	264.5	3.8	1036	1	AXO1_CHICK	P28685	gallus gall
25	264	3.8	1302	1	NRG_DROME	P20241	drosophila
26	263.5	3.8	1284	1	NRC_CHICK	P35331	gallus gall
27	261.5	3.8	1040	1	AXO1_RAT	P22063	rattus norv
28	261	3.8	858	1	NCA1_RAT	P13596	rattus norv
29	257.5	3.7	1694	1	SN_MOUSE	Q62230	mus musculus
30	257	3.7	1493	1	NEO1_MOUSE	P97798	mus musculus
31	255	3.7	1461	1	NEO1_HUMAN	Q92859	homo sapien
32	253	3.7	1271	1	MYPC_CHICK	Q90688	gallus gall
33	248.5	3.6	1451	1	MYM1_HUMAN	P52179	homo sapien

RESULT 1

KMLS_CHICK

ID KMLS_CHICK STANDARD; PRT; 1906 AA.

AC P11799; P19038;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Myosin light chain kinase, smooth muscle and non-muscle isoforms

DE (EC 2.7.1.117) (MLCK) [Contains: Telokin].

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A. (MLCK-210).

RX MEDLINE=96033976; PubMed=7589469;

RA Watterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,

RA Stepanova O.V., Shirinsky V.P.;

RT "Multiple gene products are produced from a novel protein kinase

transcription region.";

RL FEBS Lett. 373:217-220(1995).

RN [2]

RP SEQUENCE FROM N.A. (MLCK-108).

RX MEDLINE=90192792; PubMed=2315320;

RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,

RA Means A.R.;

RT "Regulatory and structural motifs of chicken gizzard myosin light

chain kinase.";

RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).

RN [3]

RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.

RX TISSUE=Fibroblast;

RC MEDLINE=90361738; PubMed=2202734;

RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,

RA Matrisian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,

RA van Eldik L.J., Watterson D.M.;

RT "Use of DNA sequence and mutant analyses and antisense

oligonucleotides to examine the molecular basis of nonmuscle

myosin light chain kinase autoinhibition, calmodulin recognition, and

activity.";

RL J. Cell Biol. 111:1107-1125(1990).

RN [4]

RP SEQUENCE OF 1259-1906 FROM N.A.

RX TISSUE=Gizzard;

RC MEDLINE=87157587; PubMed=3030394;

RA Guerrierio V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.;

RT "Domain organization of chicken gizzard myosin light chain kinase

deduced from a cloned cDNA.";

RL Biochemistry 25:8372-8381(1986).

RN [5]

RP SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).

RX TISSUE=Gizzard;

RX MEDLINE=93073972; PubMed=1444462;

RA Yoshikai S., Ikebe M.;

RT "Molecular cloning of the chicken gizzard telokin gene and cDNA.";

P13590 gallus gall
Q90610 gallus gall
O63198 rattus norv
O70468 mus musculu
Q12860 homo sapien
O60885 homo sapien
P12960 mus musculu
P16170 xenopus lae
P70478 rattus norv
Q00872 homo sapien
P32004 homo sapien
P23468 homo sapien

ALIGNMENTS

Arch. Biochem. Biophys. 299:242-247(1992).
 [6]
 RN SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
 RP MEDLINE-92236611; PubMed-1373815;
 RX Collinge M., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
 RA van Eldik L.J., Watterson D.M.;
 RA "Structure and expression of a calcium-binding protein gene contained
 RT within a calmodulin-regulated protein kinase gene.";
 RL Mol. Cell. Biol. 12:2359-2371(1992).
 CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A
 CC MYOSIN LIGHT CHAIN, WHICH LEADS TO THE FORMATION CALMODULIN/
 CC MLCK SIGNAL TRANSDUCTION COMPLEXES WHICH ALLOW SELECTIVE
 CC TRANSDUCTION OF CALCIUM SIGNALS.
 CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] -> ADP + [myosin
 CC light-chain] phosphate.
 CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; MLCK-210/non-muscle
 CC form, MLCK-108/smooth-muscle form and telokin; are
 CC produced by alternative initiation.
 CC -1- TISSUE SPECIFICITY: TELOKIN IS EXPRESSED IN GIZZARD, HEART, LUNG,
 CC INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF THE
 CC EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE GIZZARD.
 CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52876; CAA37056.1; -
 DR EMBL: X52876; CAA37057.1; -
 DR EMBL: X52876; CAA37058.1; -
 DR EMBL: M31048; AAA49069.1; -
 DR EMBL: M14953; AAA69964.1; -
 DR EMBL: M96655; AAA49083.1; -
 DR EMBL: M88283; AAA48647.1; -
 DR EMBL: M88284; AAB53768.1; -
 DR PIR: A25810; A25810.
 DR PIR: A37099; A37099.
 DR PIR: S11652; S11652.
 DR HSP: P56276; 1TLK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00041; fn3; 1.
 DR Pfam: PF00047; ig; 9.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00060; FN3; 1.
 DR SMART: SM00410; IG_like; 1.
 DR SMART: SM00408; IGC2; 8.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transferase: Serine/threonine-protein kinase; Calmodulin-binding;
 KW ATP-binding; phosphorylation; Immunoglobulin domain; Repeat;
 KW Alternative initiation.
 FT CHAIN 1 1906 MYOSIN LIGHT CHAIN KINASE, MLCK-210 (NON-
 FT MUSCLE ISOZYME).
 FT CHAIN 935 1906 MYOSIN LIGHT CHAIN KINASE, MLCK-108
 FT (SMOOTH-MUSCLE ISOZYME).
 FT CHAIN 1750 1906 TELOKIN.
 FT INIT_MET 935 935 FOR MLCK-108.
 FT INIT_MET 1750 1750 FOR TELOKIN.

FT	DOMAIN	1330	1400	FIBRONECTIN TYPE-III. PROTEIN KINASE. IG-LIKE C2-TYPE DOMAIN. ATP (BY SIMILARITY). ATP (BY SIMILARITY). BY SIMILARITY. CALMODULIN AUTOINHIBITION (AM13) REGION (POTENTIAL). CALMODULIN RECOGNITION (RS20) REGION (POTENTIAL). MOTIF IA. MOTIF IB. 4 X REPEATS, MOTIF IIA. IIA-1. IIA-2. IIA-3. IIA-4. 5 X REPEATS, MOTIF IIB. IIB-1. IIB-2. IIB-3. IIB-4. IIB-5. 4 X REPEATS, MOTIF III. III-1. III-2. III-3. III-4. CALMODULIN-BINDING. POLY-GLU. PHOSPHORYLATION. MOD_RES CONFLICT SEQUENCE	1453 1809 1459 1467 1482 1574 1716	1708 1876 1467 1482 1574 1716 1728	1730 1317 1385 1402 1833 660 676 758 774 1107 1123 1817 1833 693 1866 791 807 1140 1156 1297 1281 1851 1866 970 1226 999 1016 1061 1078 1209 1226 1700 1733 1896 1906 1748 1762 1439 1439 1906	AA; 210445 MW; AD7D8A3B69E3363 CRC64; Query Match 8.1%; Score 556.5; DB 1; Length 1906; Best Local Similarity 21.9%; Pred. No. 5.8e-19; Matches 237; Conservative 151; Mismatches 438; Indels 255; Gaps 36;
QY	267	QPPRFTQKLSREVPEGTRVQVDCIVGPPQVQVRYCEKELENSPDTHIVQAG--NLH	324					
Db	26	EAPAFPLPRNIRVQLGATARFEGKVRGYEPEQITWYRNHPLPEG-DHYVVDHSIRGIF	84					
QY	325	SLTIAEAFEDTGRYSCFASNIYGTSTSAEIVIEGVSSDSEDPNKEEMRIQKPNEV	384					
Db	85	SLVIKGVQEGDSKTYCEAANDGVQVTVVELTVEG-----NSLKYSLP	129					
QY	385	SPPTTSVAVIPPAVQAQHLVAQPRVATIQCCOSPTNYLQGLDGKPII---AAPVTKML	441					
Db	130	SSAKTPGGRL--SVPPVEH-----RPSIWGESPPKATKP	162					
QY	442	QNLASASEGQLVPECRVKGAPSKVWYREGTLIEDSPDFRILQKKPRSAEPEECTLV	501					
Db	163	NRVVREGQTRGFSCKITGRPQVTTWK-----GDHLQONERFNFETGTQYLE	214					
QY	502	IAEVFAEDSGCFTTASNKYGTVSSIAQLHVRGNEDLSNNSLSHSANSTTNLAIEPQPS	561					
Db	215	IQNVQLADAGIYCTTVNSAGKASVSAELTVQG-----PDKT	251					
QY	562	PHSEPPSVVEOPPKP-KLEGVLVNHNEPRSSSRIGLRVHP-----NLPDDKSGEASSEA	615					
Db	252	DTHAQP--LCMPKPTTLATKAIENSDFKQATNGIAKELKSTSTELMVETKDRLSAKKE	309					
QY	616	GVVTRQTTPDSXQERFNGQATKTPSPFPVKEPPVPLAKPLKLDSTQLQOLHNOVLLEQH	675					
Db	310	TFTTSREAK-DGKQGO-NOEANAVPQESRGTKGPQV-----LQKTSITILQAV	357					
QY	676	QLQNPFPSPKPEFPFXMTVLNSNAPVATTSXKQVKAPSSQTSFSLARPKYFFFTNTTAA	735					
Db	358	KAQPEPKAEPQ-----TTFIRQAE-----DKRIVQPLMTITTT	390					
QY	736	TVAPSSSPVFTLSSTPQTQRTVSKESLLVSHSPVQTKSPGGL--SIQNEPLPDPGTEPT	793					

```
Db 391 -----QENPSTGQVSPR-----SRETENRAGVRKSKVEKREP----- 424
QY 794 PPPTFTSIPSGNQFQPCVSPVPSPTSRIONPVAFSLVSSLPSPALPPTNMXKLPRSAP 853
Db 425 -----LGIPP--QFESR-----PQSLASGQELKFKSV-----SGKPPDV 460
QY 854 SMPSSQGLAKNTKSPQVNDNIR--ETKNVIRDLKGGKITTFSVDRPNQOQYKISSFEQR 911
Db 461 EWFKEGVPIKTEGIIQIYEDGTHCLWLKACLDGSG-----SYSCAAPNPR 507
QY 912 LMNEIEFRL--ERTPVDESDEIQHDEIPTGKCIAPITFDKRLKHFRTVEGSPVTFCKIV 969
Db 508 GQSTSTWLLTVKRPKEE-----VAPCFSSVLKGTCTVSEGGDFVLQCYVG 552
QY 970 GIPPKVYVFKDQKQISKRNECHKMRREGDGTCSLHTESTTDDGNYTTMAANPOGRIS 1029
Db 553 GVPVPEITWLLNEOPTOYAHSTFEA-----GVAKLTVDALPEDDGIYTCCLAENAGRAS 607
QY 1030 CSGLHMQSLPIRSLTSAGQSHRGRSVQERDKEPFRPHFLQAPCDMVVAHEGRL 1089
Db 608 CSAQVTVK-----EKKSKKAEGTQAALNKTKTFAPIFLKGLTDLKVMDSQ 653
QY 1090 CRLDCKVSGLPPELWLLNQVPLPDASHKMLVRETGVHSLLDPLTQRDAGTYKCIAT 1149
Db 654 VIMTVEVSANPCPEIILWNGKET-QETEDFHEKKGNEYSXYQVFPEDTKYTCFAW 712
QY 1150 NKTGONSFSLELSVAKEVKAPVILEKQNGCVPEGHVPRLECRVIGMPVPPVYKWKDN 1209
Db 713 NEIGETQATLTVOEPDQIQWFISKPSRVTAAGQNVLIISCAIAGDFPTVHWPKDG 772
QY 1210 ETI-PCTRERISMHQDTGYACLLIQPAKSDAGWYTLASNEAGIVSC-----T 1258
Db 773 QEITPGTGCEILQED---IFTLIRNVQSHAGQYEIQLRNQVGECSCQVSLMLRESSA 829
QY 1259 ARDIYAQWQHQPMPSPSSRYSGLSKGLDIFSASMSMESTVMYSCSSRVVVED 1318
Db 830 SRAEMLRDRES--ASSGERRDGNYGALT-----FGRTSGFKS---SSETRAABEEQ 878
QY 1319 E 1319
Db 879 E 879

RESULT 2
KMLS_HUMAN STANDARD: PRT: 1914 AA.
AC Q15746; Q14844; Q16794; Q95796; Q95797; Q95798; Q95799; Q9UBG5;
AC Q90IT9;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE (EC:2.7.1.117) (MCKK) [Contains: Telokin (Kinase related protein)
DE (KRP)].
GN MYLK OR MLCK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=97304466; PubMed=9160829;
RA Garcia J.G.N., Lazar V.L., Gilbert-Mcclain L.I., Gallagher P.J.,
RA Verin A.D.;
RT "Myosin light chain kinase in endothelium: molecular cloning and
RT regulation.";
RL An. J. Respir. Cell Mol. Biol. 16:489-494 (1997).
[2]
REVISONS.
RA Birukov K.G., Garcia J.G.N.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
DR
```

SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
TISSUE=Umbilical vein;
RX MEDLINE=99216419; PubMed=10198165;
RA Lazar V.L., Garcia J.G.N.;
RT "A single human myosin light chain kinase gene (MLCK; MYLK).";
RL Genomics 57:256-267(1999).
[4]
REVISONS (ISOFORM 2).
RA Birukov K.G., Garcia J.G.N.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE OF 923-1914 FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=96121365; PubMed=8575746;
RA Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,
RA Turnell W.G.;
RT "The human myosin light chain kinase (MLCK) from hippocampus:
RT cloning, sequencing, expression, and localization to 3cen-q21.";
RL Genomics 29:562-570(1995).
[6]
SEQUENCE OF 1614-1914 FROM N.A.
RC TISSUE=Lung, and Placenta;
RX MEDLINE=20007838; PubMed=10536370;
RA Watterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
RA Shinsky V.P., Van Eldik L.J., Haiech J.;
RT "Analysis of the kinase-related protein gene found at human chromosome
RT 3q21 in a multi-gene cluster: organization, expression, alternative
RT splicing and polymorphic marker.";
RL J. Cell. Biochem. 75:481-491(1999).
[7]
SEQUENCE OF 1456-1914 FROM N.A.
RC TISSUE=Placenta;
RA Watterson M.D.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CALCULUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
CC DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
CC EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
CC CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
CC AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
CC BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
CC IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; A NON-MUSCLE FORM (THE
CC LONGEST FORM), A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL
CC SECTION WITH NO CATALYTIC ACTIVITY); ARE PRODUCED BY ALTERNATIVE
CC INITIATION.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; 1 (SHOWN HERE), 2, 3A,
CC 3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
CC EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
CC CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
CC THE DOMINANT SPLICED VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
CC -1- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

EMBL; U48959; AAC18423.2; -.

```

DR EMBL; AF069601; AAD15921.2; -.
DR EMBL; AF069602; AAD15922.1; -.
DR EMBL; AF069603; AAD15923.1; -.
DR EMBL; AF069604; AAD15924.1; -.
DR EMBL; X85337; CAA59685.1; -.
DR EMBL; AF096771; AAD51380.1; -.
DR EMBL; AF096766; AAD51380.1; JOINED.
DR EMBL; AF096767; AAD51380.1; JOINED.
DR EMBL; AF096768; AAD51380.1; JOINED.
DR EMBL; AF096769; AAD51380.1; JOINED.
DR EMBL; AF096770; AAD51380.1; JOINED.
DR EMBL; AF096774; AAD54018.1; -.
DR EMBL; AF096771; AAD51381.1; -.
DR EMBL; AF096769; AAD51381.1; JOINED.
DR EMBL; AF096770; AAD51381.1; JOINED.
DR EMBL; X90870; CAA62378.1; -.
DR HSP; Q63450; IAO6.
DR Genew; HGNC:7590; MYLK.
DR MIM; 600922; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 8.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00410; IG_Like; 1.
DR SMART; SM00408; IGC2; 8.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative initiation; Alternative splicing.
FT CHAIN 1 1914 MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE ISOZYME.
FT CHAIN 923 1914 MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE ISOZYME.
FT CHAIN 1761 1914 TELOKIN.
FT INIT_MET 923 923 FOR MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE ISOZYME.
FT INIT_MET 1761 1761 MUSCLE ISOZYME.
FT DOMAIN 1343 1413 FOR TELOKIN.
FT DOMAIN 1464 1719 FIBRONECTIN TYPE-III.
FT DOMAIN 1711 1774 PROTEIN KINASE.
FT DOMAIN 1824 1891 CALMODULIN-BINDING.
FT NF_BIND 1470 1478 IG-LIKE C2-TYPE DOMAIN.
FT BINDING 1493 1493 ATP (BY SIMILARITY).
FT ACT_SITE 1585 1585 BY SIMILARITY.
FT DOMAIN 1906 1914 POLY-GLU.
FT DOMAIN 868 998 5 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 868 895 I-1.
FT REPEAT 896 923 I-2.
FT REPEAT 924 951 I-3.
FT REPEAT 952 979 I-4.
FT REPEAT 980 998 I-5 (INCOMPLETE).
FT DOMAIN 999 1063 6 X 12 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 999 1003 II-1 (INCOMPLETE).
FT REPEAT 1004 1015 II-2.
FT REPEAT 1016 1027 II-3.
FT REPEAT 1028 1039 II-4.
FT REPEAT 1040 1051 II-5.
FT REPEAT 1052 1063 II-6.
FT VARSPLIC 437 506 VSGIPKEVAFLEGTVPVRQESIEVEDAGSHYLCLLKA
RTDSTGTSCTASNAQGVSCSWTLQVER -> G (IN ISOFORM 2 AND ISOFORM 3B).

```

```

FT VARSPLIC 1433 1439 DEVEVSD -> MKWRCQT (IN ISOFORM 3A,
FT VARSPLIC 1473 1546 ISOFORM 3B AND ISOFORM 4).
FT VARSPLIC 1655 1705 GKFGVRLVKEKTKRWKAGFKAYSAKKNLRQISIM
FT VARSPLIC 1790 1790 NCLHHPKLVQCVDAFEKANIVMVLVSGGEL -> L
FT VARSPLIC 1933 1933 (IN ISOFORM 4).
FT CONFLICT 933 933 MISSING (IN ISOFORM 3A AND ISOFORM 3B).
FT CONFLICT 963 963 MISSING (IN ISOFORM DEL-1790).
FT CONFLICT 1022 1022 V -> M (IN REF. 5).
FT CONFLICT 1022 1022 S -> P (IN REF. 3; AAD15922).
FT CONFLICT 1048 1050 P -> A (IN REF. 5).
FT CONFLICT 1162 1162 KPM -> EAH (IN REF. 5).
FT CONFLICT 1210 1210 P -> L (IN REF. 3; AAD15922/AAD15923).
FT CONFLICT 1210 1210 L -> P (IN REF. 5).
FT CONFLICT 1280 1280 E -> D (IN REF. 3; AAD15922/AAD15923).
FT CONFLICT 1284 1284 M -> I (IN REF. 3; AAD15922/AAD15923/
FT CONFLICT 1300 1300 AAD15924).
FT CONFLICT 1316 1316 A -> G (IN REF. 5).
FT CONFLICT 1326 1326 L -> S (IN REF. 5).
FT CONFLICT 1478 1478 T -> S (IN REF. 5).
FT CONFLICT 1511 1511 V -> C (IN REF. 5).
FT CONFLICT 1563 1563 S -> T (IN REF. 3; AAD15922/AAD15923).
FT CONFLICT 1609 1609 I -> T (IN REF. 5).
FT CONFLICT 1639 1639 A -> P (IN REF. 5).
FT CONFLICT 1639 1639 G -> R (IN REF. 5).
FT CONFLICT 1639 1640 GY -> D (IN REF. 3; AAD15922/AAD15923/
FT CONFLICT 1639 1640 AAD15924).
Query Match 7.9%; Score 546; DB 1; Length 1914;
Best Local Similarity 21.4%; Pred. No. 1.8e-18;
Matches 232; Conservative 123; Mismatches 405; Indels 326; Gaps 33;
QY 264 PLGPPRFTQKLSREVPEGTROLDGIVVGPVQVWYCEGKELENSPDHITVQAG-- 321
DB 28 PLTEAPAFILPPRNLCIKEGATAKFEGRVGYPEPQVTHRWNGQPI-TSGRFLDCGIR 86
QY 322 NLHSLTIAEAEEDTGRYSCFASNIYGTDSAEIYIEGVSSSDSEGD--NKEEMNRQ 379
DB 87 GTSLVTHAVHEEDRGKYTCATNGSGARQVTVELTVEG-SFAKQLGQPVVSKTLGDRF- 144
QY 380 KPNEVSSPPTTSVIPPVPAQAHVAQPRVATIQCCQSTNYLQGLDGRPIIAPVFTK 439
DB 145 -----SASAV-----ETRPSIWGEC-----PPKPAT 165
QY 440 MLQNLASEGQVFECKRVKAGSPKPVWEYREGTLIEDSPDFRILQKKPRSMAPREICT 499
DB 166 KLGRVVVKEGQMGFESCKITGRPOQVTVLKGNVPLQ--PSARV-----SVSEKNGMQV 217
QY 500 LVIAEVAEDSGCCTCTASNKYGTVSSIAQLHVRGNEDLSNNGSLHSANSTTNLAIEPQ 559
DB 218 LEHGVNQDDVGVTCLVWNGSGKASMSAELSIQG-----LDSANRS----- 259
QY 560 PSPHSEPPSVEOPPKLEGVLVNHNEPRSSSRIGLRVHFNLPEDDKGSEASSEAGVVT 619
DB 260 -----F 260
QY 620 TROTTPDSXOERFNGQATKTPESFPVKEPPVPLAK-PKLDSTOLOQLHNOVLQHQQLQ 678
DB 261 VREIKATNSDVR-----KEVTNISKESKLDLEA-----AAK 293
QY 679 NPPSSPKPEFPFXMTVLNSNAPPVTTXKQVKAPSSQTFSLARPKYFFFTSTNTAATVA 738
DB 294 SKNCSSPQ-----RGSGPPWAANSQPQ-----P 316
QY 739 PSSSPVTLSTPOTIORTVSKESLLVSHSPVQTKSPGGLSIQNEPLPPGTEPTPPFT 798
DB 317 PRESKLESCDSRPTAPQT-----PVLQ-KTSSITLQAAVQV---EPAPRGLG 362
QY 799 FSIPTSGNQFQRCVSPVPSPTSRIQNPVAFLSVLPSLPAIPPTNAXLPRSPAPMSQ 858
DB 363 VLSPSGGERK-----RPAPPRPAT-----PPTQPGGLGSO 392
QY 859 GLAKKNTKSPQPVNDNDNIRETKNAVIDLQKKTFFSDVRPNQ---EYKISSFFQRLMNE 915
DB 859 GLAKKNTKSPQPVNDNDNIRETKNAVIDLQKKTFFSDVRPNQ---EYKISSFFQRLMNE 915

```


Db 393 DVYSKAANRRIPMEGO-----RDSAPFKPESKQSEVKEKNTQVRCFVSGIPK---PE 444
Qy 916 IEFRLEPTV--DESDEIQHD-----EIPTGKC----- 942
Db 445 VAWFLEGTVPVRQEGSIEYEDAGSHYLLKARFDSCTGTASNAOGQVSCSWTLQV 504
Qy 943 -----IAPIFDKRLKHFVRTEGSPVTFCKIVGIPVKVWFQDKQKQISKRNECHKMR 995
Db 505 ERLAVMEVAPSSVLLKDCAVIEGQDFVLQCSVRGTPVPRIITWLLNGQPIQYARSTCEA- 563
Qy 996 REGDGTCSLHISTSDDDGNTTMAANPQGRISCSGHLMVQSLPRTSRSLTSAGQSHRGR 1055
Db 564 ----GVAELHIQDALPEDGTGTYCLAENALGVQGSAAWTVH-----EKKK 605
Qy 1056 SRVQERDKPELQERRPRPFHLOAGDMVAHEGRCLRCDKCVSLPPELTWLLNGQPVLP 1115
Db 606 SRKSEVLLPVASKPTAPFLGLSLDKVMGDSQVTTVQVSGNPPPEVWILHNGNEIOE 665
Qy 1116 DASHKMLVRETGVHSLLDPLTQRDAGTYKCIATNKTKGNSFSLSVVAKEYKAPVIL 1175
Db 666 SEDFHEQRTG--QHSWLQIEVFPEDTGTTCRAWNSAGEVRTQAVLTQVDPHDGTQPVFI 724
Qy 1176 EKLONGVPEGVHVRLECRVIGMPPPVFWKDKNETIPTCTRERISMHQDTTGVACLLIOP 1235
Db 725 SKPRSVTASLGOSVLSICAIGDPFPTVHWRDGLKALCKDGTGHFEVLQNEDEVFT-LVLKK 783
Qy 1236 AKKSDAGWYTLAKNAEAGIVSCTARDIYAQWHQIP---PMSVVRP-----SG 1281
Db 784 VQVWHAQGYEILLKNRVGSCQVSLMIONSSARALPRGREPASCEDLCGGVGADGGGS 843
Qy 1282 SRYGSL 1287
Db 844 DRYGSL 849
RESULT 3
PGBL_HUMAN
ID PGBL_HUMAN STANDARD; PRT: 4393 AA.
AC P98160; Q16287;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin, and Colon;
RX MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 1018-1472 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,

RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [4]
RP SEQUENCE OF 892-1398 FROM N.A.
RC TISSUE=Fibrosarcoma;
RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC -!- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62515; CAA44373.1; -
DR EMBL; M85289; AA52700.1; -
DR EMBL; M64283; AA52699.1; -
DR EMBL; S76436; AA321121.2; -
DR EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR HSPG; P00740; 1EDM.
DR Siena-2DPAGE; P98160; -
DR Genew; HGNC:5273; HSPG2.
DR MIM; 142461; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00047; Ig; 22.
DR Pfam; PF00052; laminin_B; 3.
DR Pfam; PF00053; laminin_EGF; 7.


```
Db 3349 --NELLHFERAAPDSGR-----YCRVTNKKVGSAAFAQLLVGGPGSLPAT----- 3394
Qy 932 IQHDEIPTGKC-----IAPIDFKRLKFRVTEGSPVTCTKIVGIPVKVWFKDGOISK 987
Db 3395 -----SIPAGSTPTVVTPQLETK-----SIGASVEFHCAVPDRGTQLRWFKEGGOLPP 3444
Qy 988 RNECHKMRREGDGTCSLHISTSTSDDDGNTYTIMAANPOGRISCSGHLVQSLPIRSRLTS 1047
Db 3445 GHS-----VQDGV--LRINQLDQSCQGTVCQAHPGWKAQAQLVQALP--SVLIN 3494
Qy 1048 AGQSHGRSRVQERKDEPLQERFRPHFLQAPGDVNAHEGLRCLDKCVSLPPLPETWL 1107
Db 3495 I-----RTSVQ-----TVVGHAVEFECLALGDKPKQVTVWS 3525
Qy 1108 LNGQVPLPDASHKMLVRETGVHSLIDPLTORDACTKCIATNKTGNSFSLSVVAKE 1167
Db 3526 KVGGLHRLPG-----IVSGGV--VRIARVELADAGVRCRTATNAAGTQSHVLLVQALP 3578
Qy 1168 VKRAPVILEKLQGVGPHVRLECRVIGMPVPVFWYKK-DNETIPTCTRERISMHQDTT 1226
Db 3579 QISMP-----QEVRRVPAGSAAPPCIASGVPTPDISWSKLDGSLPPDSRLNNM----- 3627
Qy 1227 GYACLLITQAKKSDAGWYTLISAKNEAGIVSCSTARLDI 1263
Db 3628 ----LMLPSVQPDAGTYVCTATNRQGVKRAFAHLQV 3660

RESULT 4
PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Melanoma.
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule."
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RS SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan."
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
```

```
CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77174; AAA39911.1; -.
CC EMBL; J04054; AAA39899.1; -.
CC EMBL; J04055; AAA39912.1; -.
CC HSP; P01130; 1AJJ.
CC MGD; MGI:96257; Hspg2.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF_2.
CC InterPro: IPR001438; EGF_II.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003598; Ig_C2.
CC InterPro: IPR002172; LDL_recept_A.
CC InterPro: IPR000034; Laminin_B.
CC InterPro: IPR002049; Laminin_EGF.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR000082; SEA_domain.
CC Pfam; PF00008; EGF; 4.
CC Pfam; PF00047; ig; 15.
CC Pfam; PF00052; laminin_B; 3.
CC Pfam; PF00053; laminin_EGF; 8.
CC Pfam; PF00034; laminin_G; 3.
CC Pfam; PF00057; ldl_recept_a; 4.
CC Pfam; PF01390; SEA; 1.
CC PRINTS; PR0010; EGFLOOD.
CC PRODOM; PD003031; Laminin_B; 3.
CC SMART; SM00380; EGF_Lam; 7.
CC SMART; SM00001; EGF_like; 6.
CC SMART; SM00408; IGC2; 14.
CC SMART; SM00192; LDLa; 4.
CC SMART; SM00281; LamB; 3.
CC SMART; SM00282; LamG; 3.
CC SMART; SM00200; SEA; 1.
CC PROSITE; PS00022; EGF_1; 8.
CC PROSITE; PS01186; EGF_2; 5.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
CC PROSITE; PS00025; LAM_G_DOMAIN; 3.
CC PROSITE; PS01209; LDLRA_1; 4.
CC PROSITE; PS00066; LDLRA_2; 4.
CC PROSITE; PS00024; SEA; 1.
CC KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
CC Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
CC Extracellular matrix; EGF-like domain.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 3707 BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT Sulfate proteoglycan core protein.
FT DOMAIN 80 194 SEA.
FT DOMAIN 195 234 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 281 319 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 320 359 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 360 403 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 404 504 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 521 530 LAMININ EGF-LIKE 1 (N-TERMINAL).
FT DOMAIN 531 730 LAMININ DOMAIN IV 1 (DOMAIN III A).
FT DOMAIN 731 763 LAMININ EGF-LIKE 1 (C-TERMINAL).
FT DOMAIN 764 813 LAMININ EGF-LIKE 2.
FT DOMAIN 814 871 LAMININ EGF-LIKE 3.
FT DOMAIN 879 923 LAMININ EGF-LIKE 4 (INCOMPLETE).
FT DOMAIN 924 933 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 934 1125 LAMININ DOMAIN IV 2 (DOMAIN III B).
```

FT	DOMAIN	1126	1158	LAMININ EGF-LIKE 5 (C-TERMINAL).
FT	DOMAIN	1159	1208	LAMININ EGF-LIKE 6.
FT	DOMAIN	1209	1265	LAMININ EGF-LIKE 7.
FT	DOMAIN	1275	1324	LAMININ EGF-LIKE 8.
FT	DOMAIN	1325	1334	LAMININ EGF-LIKE 9 (N-TERMINAL).
FT	DOMAIN	1335	1529	LAMININ DOMAIN IV 3 (DOMAIN III C).
FT	DOMAIN	1530	1562	LAMININ EGF-LIKE 9 (C-TERMINAL).
FT	DOMAIN	1563	1612	LAMININ EGF-LIKE 10.
FT	DOMAIN	1613	1670	LAMININ EGF-LIKE 11.
FT	DOMAIN	1671	1771	IG-LIKE C2-TYPE DOMAIN 2.
FT	DOMAIN	1772	1865	IG-LIKE C2-TYPE DOMAIN 3.
FT	DOMAIN	1866	1954	IG-LIKE C2-TYPE DOMAIN 4.
FT	DOMAIN	1955	2049	IG-LIKE C2-TYPE DOMAIN 5.
FT	DOMAIN	2050	2148	IG-LIKE C2-TYPE DOMAIN 6.
FT	DOMAIN	2149	2244	IG-LIKE C2-TYPE DOMAIN 7.
FT	DOMAIN	2245	2343	IG-LIKE C2-TYPE DOMAIN 8.
FT	DOMAIN	2344	2436	IG-LIKE C2-TYPE DOMAIN 9.
FT	DOMAIN	2437	2532	IG-LIKE C2-TYPE DOMAIN 10.
FT	DOMAIN	2533	2619	IG-LIKE C2-TYPE DOMAIN 11.
FT	DOMAIN	2620	2720	IG-LIKE C2-TYPE DOMAIN 12.
FT	DOMAIN	2721	2809	IG-LIKE C2-TYPE DOMAIN 13.
FT	DOMAIN	2810	2895	IG-LIKE C2-TYPE DOMAIN 14.
FT	DOMAIN	2896	2980	IG-LIKE C2-TYPE DOMAIN 15.
FT	DOMAIN	2984	3162	LAMININ G-LIKE 1.
FT	DOMAIN	3163	3241	EGF-LIKE.
FT	DOMAIN	3245	3425	LAMININ G-LIKE 2.
FT	DOMAIN	3518	3705	LAMININ G-LIKE 3.
FT	SITE	65	67	HEPARAN SULFATE (POTENTIAL).
FT	SITE	71	73	HEPARAN SULFATE (POTENTIAL).
FT	SITE	76	78	HEPARAN SULFATE (POTENTIAL).
FT	SITE	3615	3617	MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	234	BY SIMILARITY.
FT	DISULFID	285	297	BY SIMILARITY.
FT	DISULFID	292	310	BY SIMILARITY.
FT	DISULFID	304	319	BY SIMILARITY.
FT	DISULFID	325	337	BY SIMILARITY.
FT	DISULFID	332	350	BY SIMILARITY.
FT	DISULFID	344	359	BY SIMILARITY.
FT	DISULFID	368	381	BY SIMILARITY.
FT	DISULFID	375	394	BY SIMILARITY.
FT	DISULFID	388	403	BY SIMILARITY.
FT	DISULFID	428	479	BY SIMILARITY.
FT	DISULFID	764	773	BY SIMILARITY.
FT	DISULFID	766	780	BY SIMILARITY.
FT	DISULFID	783	792	BY SIMILARITY.
FT	DISULFID	795	811	BY SIMILARITY.
FT	DISULFID	814	829	BY SIMILARITY.
FT	DISULFID	816	839	BY SIMILARITY.
FT	DISULFID	842	851	BY SIMILARITY.
FT	DISULFID	854	869	BY SIMILARITY.
FT	DISULFID	1159	1168	BY SIMILARITY.
FT	DISULFID	1161	1175	BY SIMILARITY.
FT	DISULFID	1178	1187	BY SIMILARITY.
FT	DISULFID	1190	1206	BY SIMILARITY.
FT	DISULFID	1209	1224	BY SIMILARITY.
FT	DISULFID	1211	1234	BY SIMILARITY.
FT	DISULFID	1237	1246	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1275	1287	BY SIMILARITY.
FT	DISULFID	1277	1293	BY SIMILARITY.
FT	DISULFID	1295	1304	BY SIMILARITY.
FT	DISULFID	1307	1322	BY SIMILARITY.
FT	DISULFID	1563	1572	BY SIMILARITY.
FT	DISULFID	1565	1579	BY SIMILARITY.
FT	DISULFID	1582	1591	BY SIMILARITY.
FT	DISULFID	1594	1610	BY SIMILARITY.
FT	DISULFID	1613	1628	BY SIMILARITY.
FT	DISULFID	1615	1638	BY SIMILARITY.
FT	DISULFID	1641	1650	BY SIMILARITY.
FT	DISULFID	1653	1668	BY SIMILARITY.
FT	DISULFID	1792	1839	BY SIMILARITY.
FT	DISULFID	1886	1932	BY SIMILARITY.
FT	DISULFID	1976	2021	BY SIMILARITY.
FT	DISULFID	2073	2118	BY SIMILARITY.
FT	DISULFID	2170	2215	BY SIMILARITY.
FT	DISULFID	2268	2313	BY SIMILARITY.
FT	DISULFID	2365	2413	BY SIMILARITY.
FT	DISULFID	2456	2506	BY SIMILARITY.
FT	DISULFID	2554	2599	BY SIMILARITY.
FT	DISULFID	2641	2686	BY SIMILARITY.
Query Match 5.4%; Score 375.5; DB 1; Length 3707; Best Local Similarity 19.5%; Pred. No. 4.3e-10; Matches 280; Conservative 160; Mismatches 491; Indels 507; Gaps 62				
Qy	26	HRGNERSRAEPSPNCH	-----FCSPSGA-----AEGGGQDDLPDLSAFLSQE 70	
Db	1950	HGSGPRVQVSPERTQHEGRTVRLYCRAAGVPSASITWRKEG	-----SLP----- 1996	
Qy	71	ELDESVNRLARLAINVDPLEKADETOARKRLSPDQMKHSPNLSFEPNFCQDNPRPTSKE	130	
Db	1997	-FRQAHGSRRLRLHMSVADSGEYVCRANNIDAQETSIMISVSP	-----STNS 2044	
Qy	131	SPOEAKRPQYCYTQSKKVFLNKAADFIEELSSLFKSHSKRI	-----RPRACKNHKSKLE 186	
Db	2045	PPAPASAPAPIRIESSSSRVAEGQTL	-----LNCVVPGHAAHQVTHWKRGGSLPTHQTHG 2100	
Qy	187	SQNKVMQENSSFSDLSERRERSVPIPIADTRDNEVNHAEQOEAKRREAQAAS	---E 244	
Db	2101	SRURLYQVSSADSGEYCVLSLSSGSL	-----EASVLVSITP 2137	
Qy	245	AAGDTPPGSSPSLYEPLGQPPRTQKLRSREVPEGTRVQLDLCIVVGIPPPQVRWYC	304	
Db	2138	AAANVHIPGVVP	-----PIRI--ETSSSRVAGQTL	DLSCVVPQOAHQAQVTHWK 2184
Qy	305	EGKELENSPDIIHVOAGNLHSLTIAEAFEDTGRYSCFASNIYTDSTA	----- 354	
Db	2185	RGGSLPAGHVH	-----GHMLRLNRVSPADSGEYSCVGTGSSGTLEASVLVTIEASEPS 2238	
Qy	355	-----EIIYIEGVSSDSEGD	-----PNKEEMN----- 376	
Db	2239	PIPAPGLAQPVYIESSSSHLTEGQTVDLKVCVVPQOAHQAQVTHWKRGGSLPARHGHSL	2298	
Qy	377	RIQKPNEV	-----SSPPTTSVAVIPVAPVQAQH-----LVAQPRVATQQC 416	
Db	2299	RLYQLSPADSGEYVCQVAGSSHPHEASFKLTVPSSNSSFRSLRSPVISIEPPSSTVQ	- 2357	
Qy	417	QSPTNYLQGLDGKPIIAAPVFTKMLQNLASEGOLVVFECRV	-KGAPSPKVEVYREGTLI 475	
Db	2358	-----GQDASFCKLIHEGAMPKIKVENKIRDOEL	2385	
Qy	476	ED	---SPD---FRILOKKPRMAEPEEICTLVIAEVAEFDSCGCTCTASNKYCTVSSIA 528	
Db	2386	EDNVHISPNGSIITIVAGPATM-EPT	-----ACVASNVYMAOSVW 2426	
Qy	529	QLHVRGNEDLS	-NNGSLH-SANSTTNLAAT---EPQSPPHSE---PPSVEQPPKKPLE 579	
Db	2427	NLSVHGPTVSVLPEGPHVHKMGKDITLECISGEPSSRPWTRLGIPVKLE	-----PRMF 2482	
Qy	580	GVLYNHNEPRSSRIGRLRVHFNLPEDDKGSSEAGVVTTRTPRPSXQERFNGQATKT	639	
Db	2483	GLMNSHMLKIAS	-----VKPSDAGTYVCQANALGTAQKQVELIVD-----TGT 2527	
Qy	640	PEPSFPVKEPPVPLAKPKLDSTQLQQLHNQVLLQHQHQLQ	-----NPPSSPKPEFPFX 691	
Db	2528	VAPGTP	-----QVVEESELTLAAGHTATLHCSATGNPPP----- 2562	
Qy	692	MTVLNSNAPPVATTSXQVKAP	-----SSQTESLAR-----PKYFEPSTNTTAATVA 738	
Db	2563	-----THWSKLRAPLWQHRIEGLTIVIPRVAQODSGQYICNATNSAGHTEA	2610	
Qy	739	PSSSPVFTLSTPQTQIORTVSKESLLVSHPSVQTKSPGGLSIQNEPLPPGTEPTPPPT	798	

```
Db 2611 ---TVVLHVESPYPY-----TIPEHTSAO-----PGNL-VOLQCLAHG-----TPPLT 2650
QY 799 FSIPSGNQFQRCVSPVPSPTSRIQNPVAFSLSSVPSLPAPPTNAXMLPRAPSMPQS 858
Db 2651 YQ-----NSLVGVLPKXAV-----RNQLLRLEPTVP-- 2678
QY 859 GLAKKNTKSPQVNDNIR---ETKNVIRDLGKKITFSDVRPNQOEYKISSFEQRLMNE 915
Db 2679 -----EDSGRYRCQVSNRV---GSAEFAQVL----- 2702
QY 916 IEFRLERTPVDESDDIEQHDIEPTGK---IAPIDKRLKHPRTVTEGSPVFTCKIVGI 971
Db 2703 -----VQSSSLPDTSPGGSTPTVQVTPQLETR-----NIGASVEFHCAPVNE 2747
QY 972 PVPKVYWFKDGGKQISKRNEHCKMRREGDGTCSLHIESTTSDDGNTTMAANPQGRISCS 1031
Db 2748 RGTHLRLWLEGGQLPPGHS-----VQDGV--LRIGNLQNCOGTYVCQAHGPGWQAQAT 2799
QY 1032 GHLMVQSLP-----ISRRLTSAGQSHRGRSVQERKEPLQERFRPHFLQAPGDMVAHE 1086
Db 2800 AQLIVQALPSVLINVRTSVHSV-----VVGHS 2826
QY 1087 GLRLCDCKVSGLPPELTLWLLNGQVLPDASHKMLVRETGVHSLLDPLTQRDAGTYKC 1146
Db 2827 ---VEFECLALGDKPKQVTSKVGGLHRLPG-----IVQSGT-IIRIAHVELADAGQYRC 2876
QY 1147 IAFNKTGQNSFSLSLVVA--KEVKKAPVILEKLQNGVPEGHVRLCECRVIGMPPPVFVW 1205
Db 2877 AATNAAGTTQSHVLLLVQALPQISTPPEI-----RVPAGSAAVPFCNASGYPTTATYW 2929
QY 1206 KK-DNETICTRERISHQDQTTGYACLLIQAPAKKSDAGWYTTLSAKNEAGIVSCTARLDI- 1263
Db 2930 SKVDGDLPPDSRLNNN-----LMLPSVRPEDAGTVCTATNRQGVKAFAYLVQVP 2980
QY 1264 -----YAQHHHQIOPPMVSRVSGSYGSLTSKGLDIFSAF--SSMESTWVYSCSRS 1313
Db 2981 ERVPIYPTQPSFLPLPTIKDAYRKE-----EIKITFRPDSADGMLLYNGQKRS 3030

RESULT 5
KMLS_BOVIN
ID KMLS_BOVIN STANDARD; PRT: 1176 AA.
AC Q28824;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)
DE [Contains: Telokin].
GN MLK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stomach;
RX MEDLINE=93203148; PubMed=1284247;
RA Kobayashi H., Inoue A., Mikawa T., Kuwayama H., Hotta Y., Masaki T.,
RA Ebashi S.;
RT "Isolation of cDNA for bovine stomach 155 kDa protein exhibiting
RT myosin light chain kinase activity.";
RL J. Biochem. 112:786-791(1992).
CC -1- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS OF MLCK ARE PRODUCED BY
CC ALTERNATIVE INITIATION, A NON-MUSCLE FORM (WHICH IS THE LONGEST
CC FORM); A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL SECTION WITH
```

```
CC NO CATALYTIC ACTIVITY).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S571131; AAB25794.1; -.
CC HSP; P56276; 1TLK.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003598; Ig_c2.
CC InterPro: IPR003600; Ig_like.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00047; ig; 2.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00060; FN3; 1.
CC SMART; SM00408; IGC2; 2.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;
CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
CC Alternative initiation.
CC CHAIN 1 1176 MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
CC ISOZYME.
CC INIT_MET 1022 1176
CC DOMAIN 100 291 FOR TELOKIN.
CC REPEAT 112 123 16 X 12 AA TANDEM REPEATS.
CC REPEAT 124 135 1.
CC REPEAT 136 147 2.
CC REPEAT 148 159 3 (INCOMPLETE).
CC REPEAT 160 171 4.
CC REPEAT 172 183 5.
CC REPEAT 184 195 6.
CC REPEAT 196 207 7.
CC REPEAT 208 219 8.
CC REPEAT 220 231 9.
CC REPEAT 232 243 10.
CC REPEAT 244 255 11.
CC REPEAT 256 267 12.
CC REPEAT 268 279 13.
CC REPEAT 280 291 14.
CC REPEAT 291 303 15.
CC REPEAT 303 315 16.
CC DOMAIN 603 673 FIBRONECTIN TYPE-III.
CC DOMAIN 725 980 PROTEIN KINASE.
CC DOMAIN 972 1035 CALMODULIN-BINDING.
CC NP_BIND 1084 1151 IG-LIKE C2-TYPE DOMAIN.
CC BINDING 731 739 ATP (BY SIMILARITY).
CC ACT_SITE 754 754 ATP (BY SIMILARITY).
CC ACT_SITE 846 846 BY SIMILARITY.
CC DOMAIN 1171 1176 POLY-GLU.
CC SEQUENCE 1176 AA; 128824 MW; F53DC6D4D42D4B97 CRC64;
```

```
Query Match 5.3%; Score 364; DB 1; Length 1176;
Best Local Similarity 19.8%; Pred. No. 3 9e-10;
Matches 238; Conservative 159; Mismatches 495; Indels 310; Gaps 45;
QY 22 AET-RHRGNRNRAPSSNPCHFGSPGAAEGGGODDLPDLISAFLSQBELDESVALR 80
Db 122 AETPKPLGN-----VKPAETPKPLGSKPA-----ETPKPLGSKPAETPKPLGNVK 168
```


Db 327 -----KWKFFGEGQASLSFSLHKLDEGLYTLRIVRSGGVTDHSAFLFVRDADPLVT 380
Qy 541 NG-----SLHSANSTNTLAAIEPQSPHSEPPSPVQPKLEGVLVNHNEPRSSR 593
Db 381 GAPGAPDQLCHDANRDYVITWKP-----PNTTTESP---VMGYFVDRCEVGNW 429
Qy 594 IGLRVHFNLPEDDKGSEASSAGVYVTTQTRPDSXQER---FNGQATKTPSPFPVKEPP 650
Db 430 V-----QCNDAPVKICKYPTVGLFEGRSYIFRVRVNSAGISRPSRSD 473
Qy 651 PVLAKPKLDSTQLQOLH-----NOVLLQHQLO-----NPP-----P 682
Db 474 AVAALDPLDLRLQAVHLEGEKEIAITODDLEGDAQVGPPTGVHASEISRNYVVLWSEP 533
Qy 683 SSPKEFPXMTVLSNAPPVYV---TSXKOVKAPSSQTESLARPK-YFF-----727
Db 534 PTPRGKDLMTYFIEKSVVSGTWRVNRQATVPRVAFVDMGCKSVFVRLSANRHGL 593
Qy 728 --PSTNTTA-----ATVAPSSPVFTLSS-----TPQTIQ---RTVSKESLLVSHPSVQTKS 774
Db 594 SEPSEITPIQAQDVTVVP-SAPGRVLASRNTKTSVVVQWDRPKHEEDLLGYVYDC---C 649
Qy 775 PGLLSIQNEPLPGGTEPTPPPP---TFSTPSNQ---FQPRCVSPVPSPTSRIQNPVA 828
Db 650 VAGTNLWE-----PCNHKPIGYNRFVYVHGLTGEQYIFRVKAVNAVGMSENSQESDVIK 703
Qy 829 FLSSV-LPSLP-----AIP-----PTNAXMLPR-----SA 852
Db 704 VQALTVPSHPYGTLLNCDGHSMTLWKVKPFGSGSPILGYLDKREVVHKNHNEVNSS 763
Qy 853 PSMPs-----QGLAKNNTKS-----POPVND---DN 875
Db 764 PSKPTILTVDGLTEGSLYEYFKIAAVNLAGIGEPSDSEHFKCEAWTWPPEPAYDLTFCE 823
Qy 876 IRETKNAVI-----RDLGKKTFTSDVRNQOBYKISSFEQRLMN 914
Db 824 VRDTSVLMLKAPVYSGSSPVGYFVDFREDAGEWITVDTTASRYLKVSDLQQG---K 881
Qy 915 EIEPRLERT-----PVDESDDIEQHDIEPTGKCIAP-----FDKRLKHFRVTE 958
Db 882 TYVFRVRVAVNANGVKGSDTSEPVLEAPGCTKEISAGVDQGNIIYLGFCQ-----EMTD 937
Qy 959 GSPVFTTKIVGIPVPKYWFKDGKQISKRNECHKMRREGDGTCSLHISTSDDGNYT 1018
Db 938 ASQFT-----WCKSYEEISD-DERFKIETVGDHS-KLYLKNPKDKEDLGTY 981
Qy 1019 IMAANPOG-----RISCSGHLMVQSLPIRSRL-----1045
Db 982 VVSVDTDGVSSSFVLDPEELERLMAISNEIKNPTIPLKSELAYEIFDKGRVRFWLQAEHL 1041
Qy 1046 -----TSAGQSHRGRSRVQ-1059
Db 1042 SPDASYRFINDRVSDSEIHRIKCDKATGIIEMVMDFSTENGSTYVTVIHDGAKSQS 1101
Qy 1060 -----ERDKPELQE-RFRFRPHLQAPGDMVAH-----BGRLCRLDCKVSGLPPEL-1104
Db 1102 SLVLIGDAFKTVLEAEFORKEFLKQGPHEAYELHWDVTECEVRLCKVANTKKETVF 1161
Qy 1105 TWL-----LNGQVPLDASHKMLVRETGVHSLIDPLTRQDAGTYKCIATNKTGNSPSLE 1160
Db 1162 KWLKDALYETETLPNL-----ERGICELLIPKLSKKHGEYKATLKDRGDQGVDSILE 1214
Qy 1161 LS-----VVAKEVKKAPVI-----1174
Db 1215 IAGKYDDMILAMSRVCKSAPLKVLCCTPGIRLQCFMKYFTDEMKNVCHCKDAKISS 1274
Qy 1175 -----LEKIQNC-----1181
Db 1275 EHRIGGSEEMALQICEPTKDKGKYTFEITFDGKDNHQRSLDLGQAFDEAFEFQOFK 1334
Qy 1182 -----GVP-----EGHPVRLCECRVIGMPPPVYVYKKNQNETIPTCERISM 1221

Db 1335 AAFAEAKNRGRGLIGLPDVTIMEGKTLNLTCTVFGNPDPEVIFWFKNDQDITOLS-EHFSV 1393
Qy 1222 HQDTTGYACLLIQPAKKSADAGWYLSAKNEAGIVSCSTARLDIYAQWHQIP---PPMSVR 1278
Db 1394 KVEQAKYVMTIKGVTSEDSKYSINIKNYGGEKIDVTSVY-KHGEKIPDMAPPOQAK 1452
Qy 1279 P 1279
Db 1453 P 1453
RESULT 9
DSCA_HUMAN
ID DSCA_HUMAN STANDARD; PRT; 2012 AA.
AC O60469; O60468;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Down syndrome cell adhesion molecule precursor (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
LYONS G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system";
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=20384934; PubMed=10925149;
RA Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Reinhardt R., Yaspo M.-L.,
RA Lehman H., Reinhardt R., Yaspo M.-L.;
RL "The DNA sequence of human chromosome 21";
RL Nature 405:311-319(2000).
CC -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM/CHD2-52 (SHOWN HERE)
CC AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF023450; AAC17967.1; -
 EMBL; AF023449; AAC17966.1; -
 EMBL; AF217525; AAF27525.1; -
 EMBL; AL163283; CAB90464.1; -
 EMBL; AL163282; CAB90436.1; -
 EMBL; AL163281; CAB90444.1; -
 Genew; HGNC:3039; DSCAM.
 MIM; 602523; -
 InterPro; IPR003961; FN_III.
 InterPro; IPR003962; FN_III.Repeat.
 InterPro; IPR003006; Ig_MHC.
 InterPro; IPR003598; Ig_C2.
 InterPro; IPR003600; Ig_Like.
 Pfam; PF00041; fn3; 6.
 Pfam; PF00047; Ig; 10.
 PRINTS; PR00014; ENTPEIIII.
 SMART; SM00060; FN3; 6.
 SMART; SM00410; Ig_Like; 2.
 SMART; SM00408; IgC2; 7.
 Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
 Transmembrane; Alternative splicing.
 SIGNAL 1 17 POTENTIAL
 CHAIN 18 2012 DOWN SYNDROME CELL ADHESION MOLECULE.
 DOMAIN 18 1595 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 1596 1616 POTENTIAL.
 DOMAIN 1617 2012 CYTOPLASMIC (POTENTIAL).
 DOMAIN 39 109 IG-LIKE C2-TYPE DOMAIN 1.
 DOMAIN 138 204 IG-LIKE C2-TYPE DOMAIN 2.
 DOMAIN 239 300 IG-LIKE C2-TYPE DOMAIN 3.
 DOMAIN 328 392 IG-LIKE C2-TYPE DOMAIN 4.
 DOMAIN 421 491 IG-LIKE C2-TYPE DOMAIN 5.
 DOMAIN 518 582 IG-LIKE C2-TYPE DOMAIN 6.
 DOMAIN 610 676 IG-LIKE C2-TYPE DOMAIN 7.
 DOMAIN 704 773 IG-LIKE C2-TYPE DOMAIN 8.
 DOMAIN 802 872 IG-LIKE C2-TYPE DOMAIN 9.
 DOMAIN 885 972 FIBRONECTIN TYPE-III 1.
 DOMAIN 984 1076 FIBRONECTIN TYPE-III 2.
 DOMAIN 1088 1177 FIBRONECTIN TYPE-III 3.
 DOMAIN 1189 1273 FIBRONECTIN TYPE-III 4.
 DOMAIN 1300 1366 IG-LIKE C2-TYPE DOMAIN 10.
 DOMAIN 1380 1463 FIBRONECTIN TYPE-III 5.
 DOMAIN 1477 1562 FIBRONECTIN TYPE-III 6.
 DISULFID 46 102 BY SIMILARITY.
 DISULFID 145 197 BY SIMILARITY.
 DISULFID 246 293 BY SIMILARITY.
 DISULFID 335 385 BY SIMILARITY.
 DISULFID 428 484 BY SIMILARITY.
 DISULFID 525 575 BY SIMILARITY.
 DISULFID 617 669 BY SIMILARITY.
 DISULFID 711 766 BY SIMILARITY.
 DISULFID 809 865 BY SIMILARITY.
 DISULFID 1307 1359 BY SIMILARITY.
 CARBOHYD 28 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 487 487 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 512 512 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 658 658 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 795 795 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 1142 1142 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 1160 1160 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 1250 1250 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 1271 1271 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1488 1488 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 1562 1571 NFATLAVDGS -> KEARCKEFS (IN SHORT ISOFORM).
 FT VARSPLIC 1572 2012 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1893 2012 HRPEDLIHLPPYLRMDFFLNRRGGTSDLSLQACLEPOK
 FT SRTLKRPVLEPTPEMAASASSTREGQSQWQAVATLPOR
 FT EGAELOAAKMSQSSQSLSDSRGHLKGNPNYAKSYTLV ->
 FT IGOVTSYICLIHTLEWTF (IN REF. 1).
 SQ SEQUENCE 2012 AA: 222259 MW: 0833CFB781A08334 CRC64;
 Query Match 4.6%; Score 315; DB 1; Length 2012;
 Best Local Similarity 18.2%; Pred. No. 1.5e-07;
 Matches 212; Conservative 163; Mismatches 424; Indels 364; Gaps 48;
 QY 257 SLLYYEPLQPPRFTOKLSREVPECTRVQLDCIVVGIIPPQVRWY-CEGKELENSPDI 315
 DB 22 SLLYFNASLQEVVFAST-----TGLVPCPAAGIPPVTLRWYLATGEETIDYVGI 72
 QY 316 -HIVQAGNLHSL-----TIAEAFEDTGRYSCFASNIYGTDSAEIYIEGVSSDSEGD 369
 DB 73 RHVPNGTLQIFPPPPSSFTLIHDNT--YYCTAENFSG-KIRSQDVHIAV----- 121
 QY 370 PNKEEMNRIOKNEV-----SSPPTTSVAVIPPAVQ-----AQH 403
 DB 122 -----LREPYTVRVEDOKTMRGNVAFKCIIPSSVEAYITVVSWEKDTVSLVSGSRF 173
 QY 404 LVAQPRVATQQOCS-----PNTYLOGLDGKPIIAAPVTMKLQN 443
 DB 174 LITSTGALYIKDVQNGEDGLNYRCITRHYTGETRQSARLFVSDPANSAPSLDGFHD 233
 QY 444 LSASEGOLVFECKVKGAPSPKVEWYREGTLIEDSPFRILQKKPRMAPEEICTLVIA 503
 DB 234 RKWAGQVVELPCKALGHPEPDYRWLKDNPLELSGRQ-----KTVTGLLIE 281
 QY 504 EVFAEDSGCFTCTASNKYGTVSSIAQLHVRGNEGLSNGLSHSANSTNLAIEPQPSPP 563
 DB 282 NIRPSDSGSYVCEVSNRYGTAKVIGLYVK-----KTVTGLLIE 311
 QY 564 HSEPPSVEQPPKPLEGLVYNHNEPRSSRIGLRVHFNLPEDDKGSEASSAGVVTTRQT 623
 DB 312 -----OPLKATI-----SPRK-----VKSSVGSQVSLSCSTGTGTD- 342
 QY 624 RPDSEXQERF---NGQATKTPPEPPVKEPPVLAQKLDSTQLQOLHNOVLEOHQLONP 680
 DB 343 -----QELSWRNGE-----ILNPGK--NVRITGINHENLIMDHVWKS 379
 QY 681 PPS-----SPKEFFFXMTVLNSNAPPVATTSKQVKAPSSOTFSLARPKYFFPST 730
 DB 380 GGAYQCFVRKDKLSAQD--YVQVVELEDGTPKIISAFSEKVVSPAEPVSLMCMNVK---GT 433
 QY 731 NTAATVAPSSPVFTLSSPTQRTVTSKESLLVSHPSVQTKS--PGGLSIQNEPLPPG 788
 DB 434 PLPTITWLDODDPT--LKGGSHRISQMITSEGNVYSLTSSSQVRRDGV----- 481
 QY 789 PTEPTPPPTFSIPSGNQ-----FOPKCVSIPVSPTSRIQNPVAFSLSSVLP--SLPAIP 841
 DB 482 -----YRCTANNSAGVLYQARI-----NVRGP-----ASIRPMKNITAI 517
 QY 842 PTNAMXLPVSAPMSQGLAKKNTKSPQVNDNDNIRETKNAVIRDLGKKITFSDVRPNQ 901
 DB 518 GRDPTIHC-R-VIGYPYYSIKWYKNSNLLPFNHRQVAFENNGTLK-----LSDV---QK 566
 QY 902 EYKISSFEORLMNEIEFRLERTPVDESD---DEQHDIEIPGKCI-----APIFDKRLKH 953
 DB 567 E-----VDEGEYTCNVLPQSLSTQSQVHVTVKVPPIQPFEEF 604
 QY 954 FRVTEGSPVTFCKIVGIPVP-KVYWFKDGKQIKSRNECHKMRREG-----DGPCS 1003
 DB 605 PRFSIGQRFVTPCVVVSGLDPLITTWQKDRPI-----PGSLGVITIDNIDFTSS 653
 QY 1004 LHIESTTSDDDGNYTITMAANPQGRISCSGHLMVOSLPISRLTSAGQSHRGRSRVQERDK 1063

Db 654 LRISNLSLMHNGNYTCIARNEAAVEHQSLIVR-VP----- 689
Qy 1064 EPLQERFRPHFLOAPGDWAHEGRCLRDCKVSLPPELTWLLN---GQVLPDASHK 1120
Db 690 -----PFEVQPRQDGYKAVTLNCSAEGYVPTIWKFSKAGVQFPQPIALN 740
Qy 1121 MLVRETGVHSLLDPLTQDAGTYKCIATNKTQGN-SFSLSLSVAKVKKAPVILEKLQ 1179
Db 741 GRIQVLSNGSLLIKHWVEEDSGYLVCKVSDVDGADVSKSMYLTV-----KIPAMITSYP 794
Qy 1180 NCGV-PEGHPVRLECRVGMPPPYVKKQNETIPTCTRERISMHODTTG---YACLLIQP 1235
Db 795 NTLTATQCKKEMSCATGKRPITVWEKEDRIINPEMAYLVSTKEVGEVISTLIQLP 854
Qy 1236 AKKSDAGWYTLISAKN---EAGIVSCTARLDIYAQWHHQPMPMSVSPSGSGRYSLT-SK 1290
Db 855 TVREDSGFFSCHAINSNGYEDGRIQLTVQ-----EPPDPPEIKDVKARTILRWIM 907
Qy 1291 GLDIFSASFSSMESTVYSCSSRS 1313
Db 908 GFDGNSPITGYD-----IECKNKS 926

RESULT 10
DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor suppressor protein DCC precursor.
GN DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RC STRAIN-BALB/C; TISSUE-Brain;
RX MEDLINE=96112625; PubMed=8570174;
RA Cooper H.M., Ames P., Britto J., Gad J., Wilks A.F.;
RT "Cloning of the mouse homologue of the deleted in colorectal cancer
RT gene (mDCC) and its expression in the developing mouse embryo.";
RN Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS
RC STRAIN-BALB/C; TISSUE-Brain;
RA Cooper H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: A (shown here) and B: are
CC produced by alternative initiation. A third isoform: C is
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC ISOFORM C IS EXPRESSED ONLY IN THE EMBRYO.
CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC AND REMAIN AT THIS LEVEL IN THE ADULT.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X85788; CAA59786.1; -.
DR HSP: P56276; I7LK.
DR MGD; MGI:94869; Dcc.
DR InterPro; IPR003961; FN_III_repeat.
DR InterPro; IPR003962; FN_III_repeat.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_like.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00410; IG_Like; 2.
DR SMART; SM00408; IG2; 3.
DR Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1447
FT CHAIN 85 1447
FT INIT_MET 85 85
FT DOMAIN 26 1097
FT TRANSMEM 1098 1122
FT DOMAIN 1123 1447
FT DOMAIN 154 124
FT DOMAIN 154 219
FT DOMAIN 254 317
FT DOMAIN 345 407
FT DOMAIN 426 522
FT DOMAIN 525 618
FT DOMAIN 619 716
FT DOMAIN 722 816
FT DOMAIN 840 940
FT DOMAIN 941 1042
FT DISULFID 61 117
FT DISULFID 161 212
FT DISULFID 261 310
FT DISULFID 352 400
FT CARBOHYD 60 60
FT CARBOHYD 94 94
FT CARBOHYD 299 299
FT CARBOHYD 318 318
FT CARBOHYD 478 478
FT CARBOHYD 628 628
FT CARBOHYD 702 702
FT CARBOHYD 819 838
FT VARSPLIC
SQ SEQUENCE 1447 AA; 158298 MW; 0D1F107C22D5B9F CRC64;

Query Match 4.5%; Score 307; DB 1; Length 1447;
Best Local Similarity 28.3%; Pred. No. 2.4e-07;
Matches 107; Conservative 46; Mismatches 129; Indels 96; Gaps 15;

Qy 959 GSPVTFCKIVGIPVYKFKDGKQISKNEHCKMRREGD-----TCSLHTESTSD 1012
Db 154 GDTVLLKCEVIGEMPTIHWKNOQDLN-----LPGDSRVVVLPSGALQSRLPQG 205
Qy 1013 DQGYNTTMAANPQGRSCSGHLMVLSQLPIRS-----RLTSAGOSHRSRVSQERDKEPL 1066
Db 206 DSGVYRCSARNPAS-----IRFGNEAEVRLSDPCLAR----- 238
Qy 1067 QEREFRRPHFLOAPGDWAHEGRCLRDCKVSLPPELTWLLNQPVLDPASHKMLVRET 1126
Db 239 -----QLYFLQRPNSVIAIEGKDAVLECCVGYPPSFTW-LRGEVITLRSKYSYL-L 290
Qy 1127 GVHSLLDPLTQDAGTYKCIATNKTQGN-SFSLSLSVAKVKKAPVILEKLQNCGVPEG 1186
Db 291 GGSNLLISNVTDDSGTYTCTVYTKNENISASAEITLV-----PPWFLNHPNSLYAES 345
Qy 1187 HPVLECRVIGMPPVYVKKQNE-TIPTCTRERISMHODTTGACLLIQPAKKSACGWYT 1245
Db 346 MDIEFCAGVSGKVPYTVNWKNGDVIP-----SDYFQIVGGSNLRILGVKRSDEGFYQ 399
Qy 1246 LSAKNEAGIVSCTARLDIYAQWHHQP-----PPMSVRPS-----GSRYGSLT--- 1288

```

Db 400 CVAENEAGNAQSAQL-----IVPKPAIPSSSILPSAPRDVLPVLVSSREVRLSWRP 451
QY 1289 ---SKG-LDIFSASFSSME 1302
Db 452 PBAKGNIOQTFFVFSRE 469

RESULT 11
MPSE_CHICK
ID MPSE_CHICK STANDARD; PRT: 1450 AA.
AC Q02173;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE M-protein, striated muscle.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic pectoralis muscle;
RX MEDLINE=93015907; PubMed=1400348;
RA Noguchi J., Yanagisawa M., Imamura M., Kasuya Y., Sakurai T.,
RA Tanaka T., Masaki T.;
RT "Complete primary structure and tissue expression of chicken
RT pectoralis M-protein.";
RL J. Biol. Chem. 267:20302-20310(1992).
CC -!- FUNCTION: IS A STRUCTURAL CONSTITUENT OF MYOFIBRILLAR M-BAND IN
CC STRIATED MUSCLE.
CC -!- TISSUE SPECIFICITY: IS EXPRESSED IN PECTORALIS AND CARDIAC MUSCLE.
CC -!- DEVELOPMENTAL STAGE: CAN BE DETECTED BY DAY 10-13 IN OVO. THE
CC CONTENT IS GRADUALLY INCREASED THROUGHOUT THE OVO DEVELOPMENT AND
CC REACHED ITS PEAK AFTER HATCHING.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D11474; BAA02033.1; -.
DR PIR: A44027; A44027.
DR HSP: P56276; ITLK.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FnIII_repeat.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_like.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 5.
DR PRINTS: PRO0014; FNTYPEIII.
DR SMART: SM00060; FN3; 4.
DR SMART: SM00410; IG_like; 3.
DR SMART: SM00408; IGC2; 1.
KW Muscle protein; Repeat; Immunoglobulin domain.
FT DOMAIN 142 238 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 266 364 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 368 464 FIBRONECTIN TYPE-III 1.
FT DOMAIN 496 592 FIBRONECTIN TYPE-III 2.
FT DOMAIN 597 693 FIBRONECTIN TYPE-III 3.
FT DOMAIN 696 794 FIBRONECTIN TYPE-III 4.
FT DOMAIN 797 898 FIBRONECTIN TYPE-III 5.
FT DOMAIN 899 990 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 1002 1095 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 1126 1205 IG-LIKE C2-TYPE DOMAIN 5.

```

```

FT DOMAIN 1225 1312 IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN 1333 1427 IG-LIKE C2-TYPE DOMAIN 7.
SQ SEQUENCE 1450 AA; 163400 MW; E313C7D39FE6CAC7 CRC64;

Query Match 4.48; Score 301.5; DB 1; Length 1450;
Best Local Similarity 18.6%; Pred. No. 4.3e-07;
Matches 292; Conservative 183; Mismatches 516; Indels 577; Gaps 65;

QY 143 ETOSKKVFLNKAADFEELSLSFKSH--SSKRIRPRACKNHHKSLKESQNVKQENSSSF 199
Db 31 EYARKAASQAAHY--ESTGLKGTTCRLCARRASLSA---HEAMQESRRKT-HEOKSHA 84
QY 200 SDSLERR-----ERSSVPIPIPADTRDNEY-----NHALEQOE---AKRREAQAASEAAG 247
Db 85 SD--EKRIKFASELSSL-----EREIHMARHAREOLDRLAIQRYVEENWALERH- 132
QY 248 GDTTPGSSPSLYEEPLGQPPRFTOKLRSEVPEGRVQLDICIVGIPPPQVWVWCEGK 307
Db 133 -----VVEEKISRAPETLLVRLSHRTVMEKMSVRLCFTVCGFPSPVQWY-KNE 179
QY 308 EL-----ENSPDIHIQVAGNLHSLTAAFEEDTGRYSCFASNIYGTDSAE 355
Db 180 ELITPASDPAKYSVENKYGWHV-----LHINRADFDSATYSAVATNIHQOASTNCA 231
QY 356 IYIEGVSSSDSEGDPNKEEMNRIOKPNESVSPPTSAVIPPVPAQOHLVAQPRVATIQ 415
Db 232 VVYRRPRESE-----EHPAGIMP 250
QY 416 CQSTNYLQGLDGKPIIAAPVFTKMLQNL-----ASEQLVVFECRVKGAFS-----PKVE 467
Db 251 FHLPLSY-----DVCFTDFDQVLEKFGVTEGETITLKCSSLVLTPELRLRPRAE 303
QY 468 WYREGTLIEDSPDFRILQKKPRSMAPEEICTLVIAEFAEDSCGCTCTASNYGTVSSI 527
Db 304 WYRDDVLIKDS-----KWTXLYFGEGQAALSFTHLNKDDDELGYTLRMVTKGVNecs 355
QY 528 AOLHVRGNEDLSNNG-----SLHSANSTNLAAIEPQSPHSEPPSVEQPPKPKLEG 580
Db 356 AFLVFRDADALIAAGPAMPDVKCHDANRDYVIVTWKP-----DNTTSQNP---VIG 404
QY 581 VLVNHNPRSSRIGLR--VHFNLPEDDKGSEASSEAGVVTTRQTRPDSQER---FNGQ 635
Db 405 YFVDKCE-----VGLNWMYQCN-----DAPVKICKYPTVTLGEGRSYIFRVR 446
QY 636 ATKTPSEPVKPPPPVLAKPKLDSTOLOQLH---NQVLLEQHLQNP-----PPSSPK 686
Db 447 AVNSAGISRSRSEPVAAALDPVDLERTQIVHVDGKRKIVISKDDLEDGDIQPGPTNVH 506
QY 687 EFPFXMTVLNSNAPPVATTSXKQVKAPSSQTFSLARPKYFFPSTNTTAATVAPSSS--PV 744
Db 507 ASEISKTYVVLSDPDPVPRGREPL-----TYFIEKSMVSGSWQRYNAQVAKSPRYAV 560
QY 745 FTLSSPTQTIQTVSKESLLVSHPSVQTKSPGGLSIQNEPLPGPGTEPTPPPT----- 798
Db 561 FDLAEGKPYFVYVLSANKHGIDSPSEITEPIQODIVVWPSPAPGRVATRTKTSVVVQW 620
QY 799 -----FSPISGNQFQP-----RCVSPIPV 817
Db 621 DKPKHEENLYGYIDYSVSGSNQWEPANHKPINYNRVFVHGLETFGEQYIFRKAVNAVGF 680
QY 818 SPTSRIQNPAFLSSV-LPSLP-AIPPTN-----AMXLPRSAPSMPSQGLAKNKPSP--- 868
Db 681 SENSESEAIKVAQAALTCPSYPHGICITLLNCDGHSWTLGWKAP-----KYSGSPILG 732
QY 869 -----QPVNDNIRETKNAVI-----RDLGK-----KITFSDV-----RPN--QOEYK 904
Db 733 YYIDKREANKHNWHEVNSSVISRTIYTVEDLTEDAFYEKIAAANVIGIGHPSPDSEHK 792
QY 905 ISSFEQ-----RLMNEIEFLRLTVPDE-----SDDEIQHDE 936
Db 793 CKAWTMPGPGPAYDLTVCEVRNTSLVLLWKAPVYEGKSPITGYLVLDYKEVDTEWDITANE 852
QY 937 IPTGKCIAPFDKRLKH---FRV-----TEGSPVT----- 963

```

```
Db 853 KPTSHRVKVTDLHQGHTYVFKVAVNDAGVGSSEISEPVEASPGTKEIFSGVDEEG 912
Qy 964 -----FTCKVIGIPVKVYFKDGCKQISKRNECHCKMRREGDGTCSLHSTSTSDDDGNYT 1018
Db 913 NIYLGFECK-EATDASHFLWCKGSYEEL-EDSDKFKIETKGDHS-KLYFKHPDKSDLGTYC 969
Qy 1019 IMAANPOG-----RISCSGHLMVQSLPRTSRL----- 1045
Db 970 ISVSDTDGVSFSSFLVDBEELERLMTLSNEIKNPITPLKSELAYELVDKGEVRFWIAESL 1029
Qy 1046 -----TSAGQSUR-----GRSRVQERD----- 1062
Db 1030 SPNSTYRFVINDKEVNGDRHKISCDHSNGIEMVMDKFTTIDNEGTYTVQIQDGKAKNQ 1089
Qy 1063 -----KEPQSRFR-----PHELOAPGDWAHEGLRCL--LDCKVSGGLPPP 1102
Db 1090 SLVLIGDAFKAILAESELOREFLRKQGFHSEFLYWEVTEE---CEVLLACKTANTYKKE 1146
Qy 1103 EL-TWLLNGQVPLPDASHKMLVRETGVHSLLDPLTQDAGTYKCIATNKTGQNSFLEL 1161
Db 1147 TVFKWYRNGSGIDVDEAPDL---QKGECHLTVPKLSRKDEGVYKATLSDDRGHDVSTLEL 1203
Qy 1162 S-----VVA-----KEVK----- 1169
Db 1204 SGKVNDIILALSRVSGKTASPLKILCTEGIRLQCFLKYYNEEMKVMTWSHRESKISSGE 1263
Qy 1170 -----KATL-----KA 1171
Db 1264 KMKTGGEDVAVLQITEPTEKDKGNKYFETPSDKESKRTLDLGSQAFDDALTEFFQLKA 1323
Qy 1172 PVILEKLQN---CGVP-----EGHVRLECRVIGMPPPVFVYKWKDNETIPCTRE-RISM 1221
Db 1324 AFAEKNRGKVIYGLPDVVTIMDCKTNLNCTVFGNPDPEVWFVFNKDKALENHEHYLVSL 1383
Qy 1222 HQDTGTGACLLIOPAKSDAGWYPLSAKNEAGIVSCFARLDIYAQWHH-----QIPPP 1274
Db 1384 EQGK--YASLTIKGTSEDGSKYIYVKNKYGETVDVTVSVRHGEKIPEVNOGLAKP 1441
Qy 1275 MSVRPSS 1282
Db 1442 RLIPSSS 1449

RESULT 12
DCC_HUMAN
ID DCC_HUMAN STANDARD; PRT; 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95011532; PubMed-7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RT "The DCC gene product in cellular differentiation and colorectal
RL tumorigenesis."
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE-90100559; PubMed-2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RT "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers."
```



```
QY 872 NDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISSFEQRLMNEIFRLERTPTVDESDD 931
Db 1234 -----RDEG-----FYSCSA-----RSPGLQANTS 1253
QY 932 IOHDEIPTGKCIAPFDKRLKHFVTEGSPVTTCTKIGVPPKVY-WFKDQKQISKRNE 990
Db 1254 LEURLEGVRILAP-----EAAVPEGAPITVTCADPAHAHPTLXTWYHNGRWL----- 1301
QY 991 HCKMRREGDCTSLHIBSTSDDDGNTVMAANPOGRISCGHLWQSLFIRSLTSAGQ 1050
Db 1302 -----QEGP-AASLSFVATRAAGAYSQOADAQG----- 1331
QY 1051 SHRGSRVQERDKEPLQERFRPHFLQ---APGDVMAHEGRLCR-----LDCKVSLGPP 1101
Db 1332 -----TRSSRAALQVLYAPODAVLSSFRDSRARSMAVICTVDSEPP 1374
QY 1102 PELTWLNGOPVLDPASHKMLVRETGVHSLIIDP-----LTQRD-AG--TYKC 1146
Db 1375 AELALSHDG-----KVLATSSGVHSLASGTGHVQVARNALRLQVQDVPACDDTYVC 1425
QY 1147 IATNKTGNSFSLSVVAK-EVKAPVILEKLQNCVPEGHPVRLRCRVIGMPPV--- 1202
Db 1426 TAQNLG-----SISVIGRLQVEGARVAE--PGLDVPAGALNLSCLRLGGPGPVGNS 1477
QY 1203 ---FYW---KKNETIPTCTRERISMHDQTTGYACLLIQPAKKSADAGWYTLKSAKEAGIVS 1256
Db 1478 TEAFWNRDLRLHAEPVPTLA-----FTHVARAQAGMYH 1510
QY 1257 CTARLDIYAQWHQI-----PPP-----MSVRPSGRSGRYGSL 1287
Db 1511 CLAEPLTGAASAPVMLRVLYPPKPTPTMMVFEVEPGLGRLGIL 1552

RESULT 14
MYM1_MOUSE STANDARD; PRT; 1666 AA.
AC Q62234;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myomesin 1 (SkelemIn).
GN MYOM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10J; TISSUE=Muscle;
RX MEDLINE=94012766; PubMed=8408035;
RA Price M.G., Gomer R.H.;
RT "SkelemIn, a cytoskeletal M-disc periphery protein, contains motifs
RT of adhesion/recognition and intermediate filament proteins.";
RL J. Biol. Chem. 268:21800-21810(1993).
RN [2]
RP INTERACTION WITH BETA-INTEGRIN.
RX MEDLINE=99074282; PubMed=9857037;
RA Reddy K.B., Gascard P., Price M.G., Negrescu E.V., Fox J.E.B.;
RT "Identification of an interaction between the M-band protein skelemIn
RT with beta1- and beta3-integrins. Colocalization of a skelemIn-like protein
RL J. Biol. Chem. 273:35039-35047(1998).
CC -1- FUNCTION: MAY LINK THE INTERMEDIATE FILAMENT CYTOSKELETON TO THE
CC M-DISC OF THE MYOFIBRILS IN STRIATED MUSCLE. MAY ALSO CONTACT
CC MYOSIN FILAMENTS. ALSO BINDS BETA-INTEGRINS.
CC -1- TISSUE SPECIFICITY: MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z22866; CAAB0488.1; -.
CC DR SWISS-2DPAGE; Q62234; MOUSE.
CC MGI; MGI:1341430; Myom1.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003962; FNIII_repeat.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR003598; Ig_C2.
CC InterPro; IPR003600; Ig_Like.
CC Pfam; PF00041; fn3; 5.
CC Pfam; PF00047; ig; 5.
CC PRINTS; PR00014; FNTYPEIII.
CC SMART; SM00060; FN3; 4.
CC SMART; SM00410; IG_Like; 4.
CC SMART; SM00408; IGC2; 1.
CC Immunoglobulin domain; Muscle protein; Thick filament; Repeat.
KW DOMAIN 271 334 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 398 463 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 489 575 FIBRONECTIN TYPE-III 1.
FT DOMAIN 617 703 FIBRONECTIN TYPE-III 2.
FT DOMAIN 718 802 FIBRONECTIN TYPE-III 3.
FT DOMAIN 915 1004 FIBRONECTIN TYPE-III 4.
FT DOMAIN 1020 1108 FIBRONECTIN TYPE-III 5.
FT DOMAIN 1136 1193 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 1352 1409 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 1568 1629 IG-LIKE C2-TYPE DOMAIN 5.
SQ SEQUENCE 1666 AA; 185330 MW; 6B4393AA169B2D5B CRC64;

Query Match 4.1%; Score 282.5; DB 1; Length 1666;
Best Local Similarity 19.4%; Pred. No. 4e-06;
Matches 294; Conservative 166; Mismatches 502; Indels 557; Gaps 64;

QY 96 ARKRLSPDQMKHNSPLSFEPNFCODNPSPTSSKESQO--EAKRPQYCSFQO----- 146
Db 94 SRKTASAYDYGSHGLTDSLLIED-----YSSKLSFQTKRAKVLGSEETGSLPGNLY 148
QY 147 -----KKVFLNKAADFTIEELSSLFKSHSKRIRPRACK-NHKSKLESQNKVQMNSSFS 200
Db 149 PIYSGRQVHISGIRDSEE-----RIKEAAAYIAQKTLLESEEAIAASKQSTAS 197
QY 201 DLSERRRSPVPIPADTRDNEVNHAEQEAQRRAEQAAEAAGDTPGSSPSLY 260
Db 198 QKATSKRTT-----STLQRETFEKKSRNIAIREKAEELSLKKTLEET-----QTY 244
QY 261 Y----EPLGQPPRFTOKLRSPVECTRYQLDCIVVGIPIPPQVRYWYCEGK--ELENSPD 314
Db 245 HGKLNEDHLLHAFEFITKPRSHTWKENVKLHCSVAGWPEPRITWYKNQVPIVNHANPG 304
QY 315 IHIVQAG-NLHSLTIAFAFEEDTGRYSCFASNIYGTDSIAEYIEGVSSSDSEGDPNKE 373
Db 305 KYIIESRYGMHTLEISKDFEDTAQYRASAMNVQGLSAYASVVVKRY-----KG 354
QY 374 EMNRIOQPNVSSPPTTSVAVIPPAVPOAHLVAQPRVATIQOCSPTNYLQGLDCKPIIA 433
Db 355 ELDESLLRGGVSM-PLSFVATP-----YGYASKVEIH 385
QY 434 APVFTKMLQNLSASEGOLVVFECRVKGP-----SPKVEWYRECTLTEDSPDFRLQKPP 488
Db 386 ---FDDKFDVSFGREGETMSLGCRCRVITPELKHQFQVQWVRNAPV-----SP 431
QY 489 RSMAPPE---EICTLVIAEYFAEDSGCTCTA--SNKYGTVSSIAQLHVRGNEDLSNNGS 543
Db 432 SKWQPHWSGDRATLTFSHLNKDEGLYITVRMGEYEQYS--AYVVRDAD----- 482
QY 544 LHSANSTNLAATEPQSPHSEPPSPVEQPPKPLEGVLVNHNHNSRSPSSRIGLRVHNLP 603
Db 483 -----AEIEGAPAAP-----LDVVSLDANKDY-----IISWKQP 512
QY 604 EDDKGS-----EASSEAGV- 617
```


FT	DISULFID	322	380	PROBABLE.
FT	DISULFID	422	475	PROBABLE.
FT	CARBOHYD	177	177	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	219	219	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	309	309	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	406	406	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	419	419	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	445	445	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	474	474	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	562	562	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	694	837	TLFNGLGCAIGLGVAAALLLILVYDVDSCFIRCGLLMC
FT				ITRMCGKSGSGKELEEGKAAAYKDKGKEPIVEMRTE
FT				DRITNHEDGSPVNEPNETPLTEPKLPLKEENGKEVLNA
FT				ETIEIKVSNDIIQSKEDDIKA -> NCCEANKGNGGQSWH
FT				LNAVGTFTVITWSLSCLF (IN GPI-ANCHORED
FT				ISOFORM).
SO	SEQUENCE	837 AA;	93203 MW;	70473B053A2D65A5 CRC64;
Query Match 4.0%; Score 278; DB 1; Length 837;				
Best Local Similarity 22.4%; Pred. No. 2.9e-06;				
Matches 104; Conservative 67; Mismatches 173; Indels 120; Gaps 15;				
QY	860 LAKNNTKSPQVNDNIR-----	-----ETKNA-VIRDLGKKITFSDVRPNQOEYK 904		
Db	69 LQEGVRSRLTIYNANIEDAGIYRQATDAKGOTQATVYVLEIYQKLTREVVSPOE---	125		
QY	905 ISSFEORLMNEIEFLERTP-----	-----VDESDD- 930		
Db	126 ---FKOGEDAEVVCVRSSPAPAVSWLYHNEEVTTIPDNRFAVLANNLQILNINKSDEG 182			
QY	931 -----EIOHDEIPTGKCIAPIDFKRLKHFRVT--EGSPVTFTCKIVGIPVKV 976			
Db	183 IYRCGRVEARGEIDFRDIIVVNPFAIMMPQKSFNATARGEEMTLTCKASGSPDPTI 242			
QY	977 YWFKDGKQISKNEHKMRREGDGTCSLHTESTDDNGYTTIMAAANPQGRISCSGHLMV 1036			
Db	243 SWFRNGKLI---BENEKYILKGSNT-ELTVRNLIINKGGSYVCKATNKAG-----	288		
QY	1037 QSLPIRSRLTSAGSHRGRSRVOERDKEPLQERFRPHFLQAPGDMVAHEGRLCLRLOCKV 1096			
Db	289 -----EDQKQAPLQVVFQPHILQOLKNETTSENGHV-TLVCEA 324			
QY	1097 SGLPPPELTWLLNGQPVL---PDASHKMLVRETGVH---SLIIDPLTQRDAGTYKCIATN 1150			
Db	325 EGEPVPEITWKRAIDGMFSEGDKSPDRIEVKGQHGRSSLHIRDVKLSDSGRYDCEAAS 384			
QY	1151 KTGQNSFSLSVVAKEVKAPVILEKLQNCGVPEGHVRLRCRVIGMPPPVFYWKKDNE 1210			
Db	385 RIGGHQSRMHL-----DIEYAPRFVSNQTMYSWEGNPNINISCDVTANPPASIHWRREKL 439			
QY	1211 TIPCTRERISMHQDTTGYACLL-IQPAKSDAGWYTLISAKNEAG 1253			
Db	440 LLP-AKNTTHLKHVSGRKMILEIAPTSDNDGFRYNCATNRIG 482			

Search completed: November 30, 2002, 19:05:29
Job time : 43 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:07:40 ; Search time 397 Seconds
(without alignments)
7487.753 Million cell updates/sec

Title: US-09-818-990b-2
Perfect score: 6890
Sequence: 1 MQDDSEASTSISQLRESY.....MESTWVSCSSRSVVEDEL 1320

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09818990/runat_26112002_093404_23209/app_query.fasta_1.1479
-DB=N_Geneseq_101002 -QFMF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.277 @runat_26112002_093404_23209 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEVTIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2560	37.2	2127	22	AAH13889	Human cDNA sequenc
2	2164	31.4	1228	22	AAS74798	DNA encoding novel
3	1510.5	21.9	1823	21	AAC98221	Human colon cancer
4	1421.5	20.6	4125	21	AAC76971	Human ORFX ORF2526
5	1394	20.2	2059	22	ABA06552	Human cDNA SEQ ID
6	1380	20.0	2232	22	ABA06771	Human immunoglobul
7	1133.5	16.5	702	22	ABA06731	Human cDNA SEQ ID
8	1133.5	16.5	702	22	AAS28841	Human immunoglobul
9	922.5	13.4	855	22	AAR04212	Human cDNA clone (
10	810.5	11.8	81940	24	ABK64829	Human benign prost
11	806.5	11.7	81940	22	AAS05390	Human titin (conne
12	763	11.1	790	20	AAX40026	Prostate cancer as
13	760	11.0	2244	21	AAX61262	cDNA encoding a hu
C 14	668.5	9.7	747	20	AAX40027	Prostate cancer as
15	656	9.5	392	21	AAC75084	Human ORFX ORF639
16	581.5	8.4	15272	23	ABL03697	Drosophila melanog
17	557.5	8.1	5926	24	ABN96874	Gene #3372 used to
18	557.5	8.1	5926	24	ABL61910	Colon adenocarcino
19	557.5	8.1	5926	24	ABL62450	Colon adenocarcino
20	557.5	8.1	5926	24	ABL65212	Lung cancer relate
21	557.5	8.1	5926	24	ABL69436	Prostate cancer re
22	557.5	8.1	5926	24	ABK35571	Gene MYLK differen
C 23	552.5	8.0	26097	23	ABL03696	Drosophila melanog
24	546.5	7.9	16908	24	ABQ86156	Novel human gene.
25	523.5	7.6	1768	24	AAS99901	Polynucleotide enc
26	496	7.2	20448	23	ABL19989	Drosophila melanog
27	487.5	7.1	4611	23	ABV24911	Human prostate exp
28	485.5	7.0	4611	23	ABV23407	Human prostate exp
29	485.5	7.0	4611	23	ABV29265	Human prostate exp
30	481.5	7.0	12777	23	ABL28641	Drosophila melanog
31	481.5	7.0	24789	23	ABL28640	Drosophila melanog
32	464.5	6.7	16421	23	AAS86623	DNA encoding novel
33	462.5	6.7	3910	23	AAS68120	DNA encoding novel
34	455.5	6.6	2508	22	AAS57454	Human liver cell s
C 35	441.5	6.4	24971	23	ABL19988	Drosophila melanog
36	432.5	6.3	3192	21	AAF16067	Human prostate can
37	431	6.3	14327	22	AAF54724	Nucleotide sequenc
38	429.5	6.2	14586	23	AAS87452	DNA encoding novel
39	428	6.2	3369	23	ABL19623	Drosophila melanog
40	426	6.2	23546	23	ABL02655	Drosophila melanog
41	416.5	6.0	8882	21	AAX36321	Mechanical stress
42	416.5	6.0	8883	24	AAI72583	Rat OCP cDNA. Rat
43	413.5	6.0	7679	21	AAX36322	Mechanical stress
44	408.5	5.9	8262	24	AAI72586	Human OCP cDNA. H
C 45	393.5	5.7	6192	23	ABL19622	Drosophila melanog

ALIGNMENTS

RESULT 1
AAH13889
ID AAH13889 standard; cDNA; 2127 BP.
XX
AC AAH13889;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10895.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX

PF 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 10895; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 2127 BP; 638 A; 501 C; 487 G; 501 T; 0 other;

Alignment Scores:
Pred. No.: 5,71e-120 Length: 2127
Score: 2560.00 Matches: 494
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.16% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-2 (1-1320) x AAH13889 (1-2127)

Qy 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 231 ATGCAAGACGACACATAGAGCTTCTACTTCATATCTCAGCTTCTTAAGAGAGAGCTAT 290
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 291 TTAGCTGAACACACATCGGGGAAACAAATGAGAGAGTTCGAGCGGAGCCCTCTCCAAC 350
Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu 60
Db 351 CCTTGCCATTTTCGGCAGCTCTCTTGGGGCCGCTGAAGGAGCGGAGGCGAAGATGACCTT 410
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 411 CCAGATCTTTCAGCCTTCTTGAGGCCAAGAGAAATAGACGAAAGTGTCAATTTGGCAAGA 470

Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 471 CTGGCCATCAATTACGACCCTTTGGAGAAGGCAGATCAAACTCAAGCTAGAAAAACGACTT 530
Qy 101 SerProaspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 531 TCTCCTCATCAGATGAACACACTCACCTAATTTAAGTTTGTAGCCTTAACCTTCTGCCAGGAT 590
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Db 591 AACCTCTGAAGTCCACACGCTCTAAAGAAAGCCCCAGAGGAGGCAAAAAGGCCACAGTAT 650
Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 651 TGTTCCTGAACCCAGTCCAAAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 710
Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
Db 711 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCGTGGCTGCGCAAAAC 770
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGlnLysSerSerSerPheSer 200
Db 771 CACAGAGTAAACTGGAAATCTCAAAACAAGATTATGACAGGAAACACAGCTCAGTTTCTCA 830
Qy 201 AspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
Db 831 GATCTGTCAAGAACAGACAGAAAGATCTTCTGTCCCATCCTATCCTCGGATACCAGG 890
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGlnAlaLysArgArgGluAlaGluGln 240
Db 891 GATAATCAAGTGAATACGCGCTTGGAAACAGACAGAAAGCCCAAGAGGCTGAAGCGGAGCAG 950
Qy 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Db 951 GCTGCCAGTGGCGGCTGGTGGAGACACTACACACGGGTCTTCCCTTCTATCTCTGTAC 1010
Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1011 TATGAAGAAGCTCTGGGGCAACCTCCCGGTTCCTACTCAAAAGTTTACGGAGCAGAGAAGTT 1070
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGluVal 300
Db 1071 CCAGAAGAACTCTGAGTACAGTTGGATTGCAATAGTGTAGGAATTCACCACTTCAAGTA 1130
Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
Db 1131 AGGTGGTACTGTGAAGCAAGGAGCTTGAAATTCCTCCAGATATTACATCGTCCAGCA 1190
Qy 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
Db 1191 GGAATCTGCACCTCACCTGACCATTTGCGAAGCCCTTTGAAGAGGACACAGGACGCTATTC 1250
Qy 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
Db 1251 TGTTCCTCTTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1310
Qy 361 ValSerSerSerSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
Db 1311 GTTTCCTCTTCTGACTCAGAAGGCGACCCCTAAACAAGAGAGATGAATCGAATCCAGAAG 1370
Qy 381 ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln 400
Db 1371 CCAATGAGGTGTCTATCTCTCCACCTACCTCTCGAGTCAATTCCTCCAGCAGTAGTACCCAA 1430
Qy 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
Db 1431 GCCCAGCATTTGGTGGCCCAACCTCGTGTGGCAACCATTCAGCAGTGTGAGAGCCCCACC 1490
Qy 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMet 440
Db 1491 AATTACTTGAGGAGTGGATGGAAACCTATCATTCGACGCTCTGTGTGTACAAAGATG 1550
Qy 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460

```
|||||
Db 1551 CTACAAAATTGTCAGCTTCTGAGGGTCAGCTGGTTGCTCTTTCAATGCAGAGTAAAGGA 1610
QY 461 AlaProSerProLysValGluTyrPyrArgGluGlyThrLeuIleGluAspSerProAsp 480
Db 1611 GTCCTCATCTCCTTAAGGTTGAGTGGTATAGAGAAGGACTTTAATAGAAGATTCTCCAGAT 1670
QY 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
Db 1671 TTTAGGATTTTACAGAAAAAACCTCGATCATGGCAGAGCCA 1712
RESULT 2
AAS74798
ID AAS74798 standard; cDNA; 1228 BP.
XX
AC AAS74798;
XX
DT 13-FEB-2002 (first entry)
DE
DE DNA encoding novel human diagnostic protein #10602.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR F-PSDB; ABG10611.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 10602; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1228 BP; 323 A; 319 C; 314 G; 272 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 2,46e-100 Length: 1228
Score: 2164.00 Matches: 407
Percent Similarity: 99.76% Conservative: 1
Best Local Similarity: 99.51% Mismatches: 0
Query Match: 31.41% Indels: 1
DB: 23 Gaps: 0
US-09-818-990B-2 (1-1320) x AAS74798 (1-1228)
QY 913 MetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIle 932
Db 1 ATCAATGAAATAGAGTTTCGCTTGAACGCTACTCTGTTGATGAATCAGATGAAATT 60
QY 933 GlnHisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLys 952
Db 61 CAACATGATGATGCCCGGCAAGTGTATTGCTCCCATCTTTGACAAAGACTCAAG 120
QY 953 HisPheArgValThrGluGlySerProValThrPheThrCysLysIleValGlyIlePro 972
Db 121 CACTTCGGGTACAGAAGCTCTCCAGTTACATTACCTGCAAAATTTGTTGGATACCT 180
QY 973 ValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCys 992
Db 181 GTTCCAAAGGTTTACTGGTTCAAAGATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGC 240
QY 993 LysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAsp 1012
Db 241 AAAATGAGCGAGAGAGATGGGACATGCTCTCTGCACATTGAATCCACTACGATGAT 300
QY 1013 AspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGly 1032
Db 301 GACGATGGCACTACACCATCATGGCAGCACACCCAGGGGAGAAATCACTGTTCTGGC 360
QY 1033 HisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis 1052
Db 361 CACTTGATGTACAAAGTTTGGCCATTTCGCAGTCGGCTAACCTCTGCTGCTCAGTCTCAC 420
QY 1053 ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArg 1072
Db 421 AGGGGAAGATCCCGAGTGCAGAAAGAGAGACAAAGAGAGCCCTACAGGAACGCTTTTCCGA 480
QY 1073 ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu 1092
Db 481 CCACATTTCTCGAGGCTCTCGGGATATGCTAGCTCATGAGGGGCCCTCTGTGGCTG 540
QY 1093 AspCysLysValSerGlyLeuProProGlu-LeuThrTrpLeuLeuAsnGlyGlnPr 1112
Db 541 GACTGTAAGGTGAGTGGTTTACCGCCCGGAGCATGACATGGCTACTCAATGGCCAACC 600
QY 1112 oValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLe 1132
Db 601 TGTGCTACCAAGATGCTCCACAAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCT 660
QY 1132 uIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysTh 1152
Db 661 CATTGACCCCTCACTCAGCGCGCAGCGGAGCTATAAGTGCATCGCTACCAACAAAC 720
QY 1152 rGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaPr 1172
Db 721 CGGCGAGAATTCITTTAGTCTGGAGCTCTCTGTAGTAGCCCAAGAGGTGAAGAAAGCACC 780
QY 1172 oValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuG 1192
Db 781 TGTGATCTCTGGAGAAATACAGAACTGCGGTGTTCCGGAAGGCCACCCCGTGGAGACTGGA 840
QY 1192 uCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrI 1212
Db 841 GTCCCGGCTGATAGGCATGCCACCTGTGTTCTACTGGAAGAAAGACAAATGAGACAT 900
QY 1212 eProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLe 1232
Db 901 CCCTTGCAACAGAGAGAGATCAGTATGCACCAGGACACAAACAGGGGTATGCTGCTCTCT 960
```


Qy	921	GluArgThrProValAspGluSerAspGluLeuGlnHisAspGluLeuProThrGly	940
Db	583	GAAGGTCCTCCTGGATGAATCAGGTATGAGATGTCCTGTGGAA	642
Qy	941	LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer	960
Db	643	AACGGAAATGGACCATCTTTGAGATGAAGCTGAACCATTTACAAGATCTTTTGAGGGGAATG	702
Qy	961	ProValThrPheThrCysLysIleValGluIleProValProLysValTyrTrpPheLys	980
Db	703	CCAGTAACTTTTCATGTAGAGTGGCTGGAATCTCAAAGCCCAAGATCTATTGGTTTAAA	762
Qy	981	AspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGly	1000
Db	763	GATGGGAAGCAGATCTCTCCAAAGAGTGATCACTACACCATTTCAAAGAGATCTCGATGGG	822
Qy	1001	ThrCysSerLeuHisIleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMet	1020
Db	823	ACCTGCTCCCTCCATACCACAGCCTCCACCCCTAGATGATGATGGCAATATTACAATTTATG	882
Qy	1021	AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro	1040
Db	883	GCTGCAAAACCTCAGGCGCCATCACTGTTACTGACGGCTAATGGTACAGCGTGTCCAAC	942
Qy	1041	IleArgSerArgLeuThrSerAlaGlyGlnSerHis-----ArgGly	1054
Db	943	CAAAGAGGTGGAAGTCCCGGCTCCCTCAGGCCATCTCATGTGCAAGAGCCCTCGTTCT	1002
Qy	1055	ArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHis	1074
Db	1003	AGATCAAGGACAGTGGAGCACAAAATGAACCAATTACAGAGCGATTTCTCAGACCTCAC	1062
Qy	1075	PheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCys	1094
Db	1063	TTCTTTGAGGCTCCTGGAGATCTGACTGTTCAAGAAAGAAACACTCTCGCAAGTAATGGATCGC	1122
Qy	1095	LysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeu	1114
Db	1123	AAAGTCAGTGGGTTACCAACCCAGATCTAAAGCTGCACTAGATGGAAGCCCGTACGC	1182
Qy	1115	ProAspLaserHisMetLeuValArgGluThrGlyValHisSerLeuLeuIleAsp	1134
Db	1183	CCTGACAGTCTCCAAAGATGCTGGTGGCTGAGAACGGGGTGCACTCTCTGATCATAGAG	1242
Qy	1135	ProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGln	1154
Db	1243	CCAGTCAGTCAGCTGATCGCGGCATCTACACATGTATAGTACCACCGAGAGGACAG	1302
Qy	1155	AsnSerPheSerLeuGluLeuSerValAlaLysGluValLysLysAlaProValIle	1174
Db	1303	AACTCAATTCAGCTGGAGCTTTGGTGGTCTGCTAAAGAAGCACACAACCCCGTGTGT	1362
Qy	1175	LeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArg	1194
Db	1363	ATTGAGAAGCTCCAAACACAGAGTTGCTGTAGTGGTACCAGCTGCGGCTGGAATGTCTGT	1422
Qy	1195	ValIleGlyMetProProProValPheTyrTrpLysLysAspAsnGluThrIleProCys	1214
Db	1423	GTATTTGGAGTGGCCACCTCAGATATTTTGGAGAAGAAATGAATCACTCACTCAC	1482
Qy	1215	ThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGln	1234
Db	1483	AGCACTACCCGATGAGCATGCACCAGGACACACCGGCTACATCTGCCGTCTCATTCAG	1542
Qy	1235	ProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle	1254
Db	1543	GGAGCCCAAAAGAAGATGCTGGTGGTATACTGTGTGAGCCCAAGATGAAGCAGGGATT	1602
Qy	1255	ValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis-----GlnIlePro	1272
Db	1603	GTGTCTGTACTGCCAGGTGGACGTTTTACACCCAGTGGCATCGACGAGTCACAGACACC	1662
Qy	1273	ProPrometSerValArgProSerGlySerArgTyrGlyValSerLeuThrSerLysGlyLeu	1292

D	b	1663	AAGCCAAAAAAGTACGGCCTCAGCCAGTCGCTATGCAGCACTTTTCGGACCAGGACTA	17293
Q	y	1293	AspilePheSerAlaPheSerSerMetGluSerThrMetValTyrSerCysSerSerArg	1312
D	b	1723	GACATCAAACGACGGTTCACAACCT---GAGGCCAACCCCATCTCACCTGACACTGAATACT	1779
Q	y	1313	SerValValGluSerAspGluLeu	1320
D	b	1780	GCCTTGGTAGAAAAGTAGGACCTG	1803
R	E	S	U	L
I	D	ABA06552	standard; cDNA; 2059 Bp.	
X	X	ABA06552;		
X	X	XX		
D	T	XX		
D	E	XX	10-JAN-2002 (first entry)	
X	X	XX	Human cDNA SEQ ID NO: 218.	
K	W	XX	Human; gene therapy; neural disorder; immune system disorder;	
K	W	XX	muscular disorder; reproductive disorder; gastrointestinal disorder;	
K	W	XX	pulmonary disorder; cardiovascular disorder; renal disorder;	
X	X	XX	proliferative disorder; inflammation; ss.	
O	S	XX	Homo sapiens.	
X	X	XX	WO200154474-A2.	
P	N	XX	02-AUG-2001.	
P	D	XX	17-JAN-2001; 2001WO-US01349.	
P	F	XX	31-JAN-2000; 2000US-179065P.	
P	R	XX	04-FEB-2000; 2000US-180628P.	
P	R	XX	24-FEB-2000; 2000US-184664P.	
P	R	XX	02-MAR-2000; 2000US-186350P.	
P	R	XX	16-MAR-2000; 2000US-189874P.	
P	R	XX	17-MAR-2000; 2000US-190076P.	
P	R	XX	18-APR-2000; 2000US-198123P.	
P	R	XX	19-MAY-2000; 2000US-205515P.	
P	R	XX	07-JUN-2000; 2000US-209467P.	
P	R	XX	28-JUN-2000; 2000US-214886P.	
P	R	XX	30-JUN-2000; 2000US-215135P.	
P	R	XX	07-JUL-2000; 2000US-216647P.	
P	R	XX	07-JUL-2000; 2000US-216880P.	
P	R	XX	11-JUL-2000; 2000US-217487P.	
P	R	XX	11-JUL-2000; 2000US-217496P.	
P	R	XX	14-JUL-2000; 2000US-218290P.	
P	R	XX	26-JUL-2000; 2000US-220963P.	
P	R	XX	26-JUL-2000; 2000US-220964P.	
P	R	XX	14-AUG-2000; 2000US-224518P.	
P	R	XX	14-AUG-2000; 2000US-224519P.	
P	R	XX	14-AUG-2000; 2000US-225213P.	
P	R	XX	14-AUG-2000; 2000US-225214P.	
P	R	XX	14-AUG-2000; 2000US-225266P.	
P	R	XX	14-AUG-2000; 2000US-225267P.	
P	R	XX	14-AUG-2000; 2000US-225268P.	
P	R	XX	14-AUG-2000; 2000US-225270P.	
P	R	XX	14-AUG-2000; 2000US-225447P.	
P	R	XX	14-AUG-2000; 2000US-225757P.	
P	R	XX	14-AUG-2000; 2000US-225758P.	
P	R	XX	14-AUG-2000; 2000US-225759P.	
P	R	XX	18-AUG-2000; 2000US-226279P.	
P	R	XX	22-AUG-2000; 2000US-226681P.	
P	R	XX	22-AUG-2000; 2000US-226868P.	
P	R	XX	22-AUG-2000; 2000US-227182P.	
P	R	XX	23-AUG-2000; 2000US-227009P.	
P	R	XX	30-AUG-2000; 2000US-228924P.	
P	R	XX	01-SEP-2000; 2000US-229287P.	
P	R	XX	01-SEP-2000; 2000US-229343P.	
P	R	XX	01-SEP-2000; 2000US-229344P.	

```
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.

PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
XX PI
XX P-PSDB; ABB10330.
XX
XX WPI; 2001-476161/51.
XX P-PSDB; ABB10330.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
XX polypeptide is used in preventing, treating or ameliorating a medical
XX condition
XX
XX Claim 1; SEQ ID NO: 218; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
XX DNAs. These can be used in the treatment of neural, immune system,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal and proliferative disorders and inflammation. The present sequence
XX is a cDNA of the invention.
XX
XX SQ Sequence 2059 BP; 589 A; 460 C; 502 G; 508 T; 0 other;

Alignment Scores:
Pred. No.: 2.08e-61 Length: 2059
Score: 1394.00 Matches: 261
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.62% Mismatches: 0
Query Match: 20.23% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-2 (1-1320) x ABA06552 (1-2059)
Qy 1059 GlnGluArgaspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAla 1078
Db 7 AGAGAAAGAGACAAAGAGCCCTACAGGACGCTTTTCCGACACACATTTCTCGAGGCT 66
Qy 1079 ProGlyAspMetValAlaHisGluGlyArgLeuGlyArgLeuGlyArgLeuGlyArgLeuGly 1098
Db 67 CCTGGGATATGGTAGCTCATGAGGGGGCCCTCTGTGGCTGGACTGTAAAGTGAGTGGT 126
Qy 1099 LeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSer 1118
Db 127 TTACCGCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTCTACCATGAGTCCCTCC 186
Qy 1119 HisLysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuLeuLeuLeuLeuLeu 1138
Db 187 CACAAGATGCTGGTCAGGGAGCGGAGTCCCACTCTCTGCTCAATGACCCACTCACTCAG 246
Qy 1139 ArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSer 1158
```

```
Db 247 CGCGACGCGAGGACCTATAAGTCGATCGCTACCAACAAACCGGCGCAAAATCTCTTTAGT 306
Qy 1159 LeuGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeu 1178
Db 307 CTGGAGCTCTCTGTAGTAGCAAGAGGTGAAGAAACACCTGTGATCCTCGGAAACTA 366
Qy 1179 GlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMet 1198
Db 367 CAGAACTGCGGTGTTCGGAAGGCCACCCCGTGAGACTGGAGTCCGCGTGATAGGCATG 426
Qy 1199 ProProProValPheTyrTrpPlyLysAspAsnGluThrIleProCysThrArgGluArg 1218
Db 427 CCCCACCTGTGTCTACTGGAAGAAACAATGAGACCATCCCTTGCCACGAGAGAGG 486
Qy 1219 IleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLys 1238
Db 487 ATCAGTATGCACCAAGGACACAAACAGGGTATGCCCTTCTCATTCAGCCAGCCCAAGAA 546
Qy 1239 SerAspAlaGlyTyrTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThr 1258
Db 547 TCACAGCGTGGATGGTACACGCTTGTACCCCAAGAAATGNAAGCCGCATCGTCTGTGCACT 606
Qy 1259 AlaArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProPrometSerValArg 1278
Db 607 GCCAGGCTGGATATATACGCTCAGTGGCACCACCATCAGATCCCAACGCCCATCTCTGTCCGG 666
Qy 1279 ProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPhe 1298
Db 667 CCCAGTGGCAGTCGCTACGGATCTCTCACCAGTAAGGACTTGACATATTTTCGCCCTTT 726
Qy 1299 SerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAsp 1318
Db 727 TCCTCCATGGAAACGACGACGTCGTATTTCATGCTCTCTCGGAGTGTAGTCGACAGTGTAT 786
Qy 1319 GluLeu 1320
Db 787 GAACCT 792

RESULT 6
AAS28777
ID AAS28777 standard; cDNA; 2232 BP.
XX
AC AAS28777;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immunoglobulin encoding cDNA SEQ ID No 23.
XX
KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN W0200155315-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01326.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225368.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
```

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Barash SC, Ruben SM;

WPI: 2001-457725/49.

P-PSDB; AAU17989.

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis -

Claim 1; SEQ ID No 23; 551pp; English.

XX Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the immunoglobulin polypeptides of the invention, and primers for the polynucleotides. The polynucleotides and polypeptides can be used to diagnose a pathological condition or a susceptibility to a pathological condition in a subject by determining the presence or absence of a mutation in a DNA sequence or determining the presence or amount of expression of the protein. Alternatively the identification of a binding

CC partner to a sequence allows determination of changes in protein activity. The sequences can be used as research tools for receptors or other signal transduction pathway proteins that interact with the polypeptides of the invention and can be used to treat, prevent or diagnose various types of disorders such as neurological disorders, cardiovascular disorders, gastrointestinal disorders, reproductive disorders, immune system disorders, renal disorders, muscular disorders, pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2232 BP; 617 A; 508 C; 528 G; 569 T; 10 other;

Alignment Scores:

Pred. No.:	1.15e-60	Length:	2232
Score:	1380.00	Matches:	239
Percent Similarity:	99.23%	Conservative:	0
Best Local Similarity:	99.23%	Mismatches:	2
Query Match:	20.03%	Indels:	0
DB:	22	Gaps:	0

US-09-818-990B-2 (1-1320) x AAS28777 (1-2232)

QY	1060	GluArgAspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaPro	1079
DB	3	GRAAGAGACAAAGAGCCCTACAGAAACGCTTTTCGGACCACATTTCTCGAGGCTCCT	62
QY	1080	GlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeu	1099
DB	63	GGGGATATGGTAGTCATGAGGGCGCCTCTGTCGGCTGGACTGTAAGGTGAGTGGTTTA	122
QY	1100	ProProGluLeuThrTrpLeuLeuAsnGlyGluProValLeuProAlaSerHis	1119
DB	123	CCGCCCCCGAGCTGACATGGCTACTCAATGGCCACCTGTGTACAGATGCTCCAC	182
QY	1120	LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArg	1139
DB	183	AAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTATTGACCCACTCCTCAGGCG	242
QY	1140	AspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu	1159
DB	243	GACGACGGGACCTATAAGTGCATCGCTACCACAAACCCGGCAGAAATCTTTAGTCTG	302
QY	1160	GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGlnGln	1179
DB	303	GAGCTCTGTAGTAGCAAGAGGTGAAGAAAGCACCTGTGATCTCTGGAGAACTACAG	362
QY	1180	AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro	1199
DB	363	AACTGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGCATGCC	422
QY	1200	ProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle	1219
DB	423	CCACCTGTGTCTACTGGAGAAAGACATGAGACCATCCCTTTCACCAAGAGAGAGATC	482
QY	1220	SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSer	1239
DB	483	AGTATGACACGAGACACACAGGGTATGCTCTCTCTCATTCAGCCAGCCAGAAAWCA	542
QY	1240	AspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla	1259
DB	543	GACGCTGGATATACGCTAGTGGCCACCATCAGATCCACCGSCCATGTCTGTCCGCC	602
QY	1260	ArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProMetSerValArgPro	1279
DB	603	AGCTGGATATACGCTAGTGGCCACCATCAGATCCACCGSCCATGTCTGTCCGCC	662
QY	1280	SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSer	1299
DB	663	AGTGGGAGTCCGCTACGGATCTCTCACCAGTAAAGAGGTTGACATATTTCTGCTTTTC	722
QY	1300	SerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAspGlu	1319

|||||
Db 723 TCCATGAAACGACGAGTGCTATTTCATGCTCTCTCGGAGTGTAGTGAGAGTGATGAA 782
Qy 1320 Leu 1320
|||
Db 783 CTT 785
RESULT 7
ABA06731
ID ABA06731 standard; cDNA; 702 BP.
XX
AC ABA06731;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 397.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226686P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 05-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.

```
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR P-PSDB; ABB10509.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 1; SEQ ID NO: 397; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
XX Sequence 702 BP; 179 A; 195 C; 187 G; 137 T; 4 other;
SQ
Alignment Scores:
Pred. No.: 7.85e-49 Length: 702
Score: 1133.50 Matches: 226
Percent Similarity: 96.17% Conservative: 0
Best Local Similarity: 96.17% Mismatches: 9
Query Match: 16.45% Indels: 4
DB: 22 Gaps: 0
US-09-818-990B-2 (1-1320) x ABA06731 (1-702)
QY 1060 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 1079
Db 3 GAAAGAGACAAAGAGCCCTACAGAACGCTTTTCCGACACACATTTCTCGAGGCTCCT 62
QY 1080 GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeu 1099
Db 63 GGGGATATGGTAGCTCATAGAGGGCGCCTCTGTGGCTGGACTGTAAGGTGAGTGGTTA 122
QY 1100 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 1119
Db 123 CG-CCCCGGAGCTGACATGGCTACTCAATGGCCAACTCTGTCTACCATGCTCCAC 181
QY 1120 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArg 1139
Db 182 AAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCGC 241
QY 1140 AspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 1159
Db 242 GACGA-GGGACCATATAAGTCATCGCTTACCAACAAACCGGGCAGAAATCTTTAGTCTG 300
QY 1160 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGlnLysLeuGln 1179
Db 301 GAGCTCTCTAGTAGCCAAAGAGGTGAAGAAAGACACCTGTGATCCTGGAGAAACTACAG 360
QY 1180 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 1199
Db 361 AACTGGCGGTGTTCCGAAAGGCCACCCCGTGGAGTGGAGTGGCGGTGATAGCATGCC 420
QY 1200 ProProValPheTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle 1219
Db 421 CCACCTGTGTTCTACTGGAAGAAAGACATGACACCATCCCTTGCCACCAAGAGAGATC 480
QY 1220 SerMetHisGlnAspThrThrGlyTyraAlaCysLeuLeuLeuGlnProAlaLysLysSer 1239
Db 481 AGTATGCACCCAGGACACACAGGATGCTGCTCTTCATTGACCGCAGCAAGAAATCA 540
QY 1240 AspAlaGlyThrTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 1259
Db 541 GACGCTGGATGGTACACGTTGTGAGCCCAAGAAATGAAGCCGCGCATCGNGTCTGCTGCACTGCC 600
QY 1260 ArgLeuAspIleTyraAlaGlnTrpHisHisGlnIleProProProMetSerValArgPro 1279
Db 601 AGGCTGGATATATACGCTCAGTGGCACCATCAGATCCACSGCCATGTC--TGCCGGGCC 658
QY 1280 SerGlySerArgTyrglySerLeuThrSerLysGlyLeuAspIle 1294
Db 659 AGNGGCAGTCGNTACGGATCTCTAACAGT-AAAGGACTTGACATA 702
RESULT 8
AAS28841
ID AAS28841 standard; cDNA; 702 BP.
XX
XX AAS28841;
XX
XX 07-NOV-2001 (first entry)
XX Human immunoglobulin encoding cDNA SEQ ID No 87.
XX
XX Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
XX Homo sapiens.
XX
XX WO200155315-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01326.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
```

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231142.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231144.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231141.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457725/49.

P-PSDB; AAU18053.

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis -

Claim 1; SEQ ID No 87; 551pp; English.

Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the immunoglobulin polypeptides of the invention, and primers for the polynucleotides. The polynucleotides and polypeptides can be used to diagnose a pathological condition or a susceptibility to a pathological condition in a subject by determining the presence or absence of a mutation in a DNA sequence or determining the presence or amount of expression of the protein. Alternatively the identification of a binding partner to a sequence allows determination of changes in protein activity. The sequences can be used as research tools for receptors or other signal transduction pathway proteins that interact with the polypeptides of the invention and can be used to treat, prevent or diagnose various types of disorders such as neurological disorders, cardiovascular disorders, gastrointestinal disorders, reproductive disorders, immune system disorders, renal disorders, muscular disorders, pulmonary disorders, proliferative disorders and cancer.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 702 BP; 179 A; 195 C; 187 G; 137 T; 4 other;

Alignment Scores:

Pred. No.: 7, 85e-49 Length: 702
Score: 1133.50 Matches: 226
Percent Similarity: 96.17% Conservativity: 0
Best Local Similarity: 96.17% Mismatches: 9
Query Match: 16.45% Indels: 4
DB: 22 Gaps: 0

US-09-818-990B-2 (1-1320) x AAS28841 (1-702)

QY 1060 GluArgAspLysGluProLeuGlnGluAurPhePheArgProHisPheLeuGlnAlaPro 1079
DB 3 GAAAGACAAAGAGCCCTACAGAACGCTTTTCCGACACATTTCTCGAGGCTCT 62
QY 1080 GlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeu 1099
DB 63 GGGGATATGGTAGCTCATGAGGGGCGCTCTGTGGGCTGGACTGTAAGGTGAGTGGTTTA 122
QY 1100 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 1119
DB 123 CG-CCCCGGAGTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCCCTCCAC 181
QY 1120 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArg 1139
DB 182 AAGATGCTGTGAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCGC 241
QY 1140 AspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 1159
DB 242 GACGA-GGGACCTATAAGTGCATGCTACCAACAAACCGGGCAGAAATCTTTTAGTCTG 300
QY 1160 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 1179
DB 301 GAGCTCTGTAGTAGCAAGAGGTGAAGAACCACTGTGTATCTGGAGAACTACAG 360
QY 1180 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 1199
DB 361 AACTGCGGTGTCCGGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCC 420
QY 1200 ProProValPheTrpTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle 1219
DB 421 CCACCTGTGTCTACTGGAGAAAGCAATGAGACCATCCCTTGCCACAGAGAGATC 480
QY 1220 SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSer 1239
DB 481 AGTATGCCAGGACACACAGGGTATGCTGCTCTCTCATTCAGCCAGCCCAAGATCA 540
QY 1240 AspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 1259
DB 541 GACGCTGGATGGTACAGGTTGTACCAAGAAATGAAGCCGCATCGNGTCTGCACTGCC 600
QY 1260 ArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProMetSerValArgPro 1279
DB 601 AGGCTGGATATATACGCTCACTAGTGCCACCATCAGATCCACAGCGGCATGTC--TGCCGGCC 658
QY 1280 SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIle 1294
DB 659 AGNGCGCATCGNTACGATCTCTAACAGT-AAAGACTTGACATA 702

RESULT 9
AAH04212

ID AAH04212 standard; cDNA; 855 BP.

XX AC AAH04212;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA clone (5'-primer) SEQ ID NO:1047.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX EP1074617-A2.
PN 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 1; SEQ ID 1047; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences. AA892446 to
CC AA895893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 855 BP; 257 A; 214 C; 176 G; 205 T; 3 other;
Alignment Scores:
Pred. No.: 3, 88e-38 Length: 855
Score: 922.50 Matches: 185
Percent Similarity: 95.45% Conservativity: 4
Best Local Similarity: 93.43% Mismatches: 9
Query Match: 13.39% Indels: 1
DB: 22 Gaps: 0
US-09-818-990B-2 (1-1320) x AAH04212 (1-855)
QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
DB 231 ATGCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 290
QY 21 LeuAlaGluThrArgHisArgGlyAsnGluArgSerArgAlaGluProSerSerAsn 40
DB 291 TTACGTGAACACACACATCGGGGAAACAATGAGAGGAGTCCGAGCGGCCCTCTCCCAAC 350
QY 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu 60
DB 351 CTTTGCCATTTTCGGCAGTCTCTGGGCGCGCTGAAGGAGCGGAGGCGCAAGATGACCTT 410

Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
 Db 411 CCAGATCTTTAGCCCTTCTGAGCCAAAGAAGAAATTAGACGAAAGTGTCAATTGGCAAGA 470

Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
 Db 471 CTGGCCATCAATTAGACCCCTTTGGAGAGGCGAGATGAAGCTCAAGCTAGAAAAGACTT 530

Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
 Db 531 TCTCTGATCAGATGAACACTCACCTAATTAAGTTTGGACCTTAACCTTCTGCCAGAT 590

Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
 Db 591 AACCTCGAAGTCCACCAAGCTCTAAAGAAAGCCCGAGGCGGCAAAAGGCCACAGTAT 650

Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
 Db 651 TGGTCTGAACCCAGTCCAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 710

Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
 Db 711 CTATCTCTCT--TTCAATCCACAGCTTCCAAAGGATTAGACCTCGTGCCTGCAGAAAC 768

Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSer 198
 Db 769 CNCAGAGTNAACTGGAATCTCAAAACAAAGTATGCAGGAAACAGCTCAGTTCT 822

RESULT 10
 ABK64829
 ID ABK64829 standard; DNA; 81940 BP.

XX AC ABK64829;
 XX DT 18-JUN-2002 (first entry)
 XX DE Human benign prostatic hyperplasia gene #724.
 XX KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
 XX OS Homo sapiens.
 XX PN WO200212440-A2.
 XX PD 14-FEB-2002.
 XX PF 07-AUG-2001; 2001WO-US24708.
 XX PR 07-AUG-2000; 2000US-22323P.
 XX PR 05-JUN-2001; 2001US-0873319.
 XX PA (GENE-) GENE LOGIC INC.
 XX PA (NTSB) JAPAN TOBACCO INC.
 XX PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
 XX WPI; 2002-257476/30.
 XX PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
 PT detecting expression levels of one or more genes in prostate cells from
 PT patient that are differentially regulated compared to normal prostate
 PT cells -
 XX PS Disclosure; Page 405-429; 444pp; English.
 XX CC The invention relates to a method of diagnosing (I) the onset or
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
 CC or identifying an agent that modulates the onset or progression of BPH.
 CC The method is based on changes in gene expression in BPH tissue isolated
 CC from patients exhibiting different clinical states of prostate
 CC hyperplasia as compared to normal prostate tissue. (I) comprises
 CC detecting the expression levels of one or more genes in prostate cells

CC from the subject that are differentially regulated compared to normal
 CC prostate cells. (II) comprises preparing a first gene expression profile
 CC of BPH cells or BPH-like cell population, exposing the cells to the
 CC agent, preparing a second gene expression profile of the agent exposed
 CC cells, and comparing the first and second gene expression profiles.
 CC (I) is useful for diagnosing the onset or progression of BPH. (II) is
 CC useful for identifying an agent that modulates the onset or progression
 CC of BPH. The methods are useful to present information identifying
 CC the expression level in a tissue or cells, by comparing the expression
 CC level of genes given in the specification in the tissue or cells to the
 CC level of expression of gene in the database, and displaying the
 CC expression levels of at least one gene in the tissue or cell sample
 CC compared to the expression level in BPH. Agents using (II) are useful for
 CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
 CC benign prostatic hyperplasia gene sequences of the invention.

XX Sequence 81940 BP; 26373 A; 17097 C; 18901 G; 19569 T; 0 other;

Alignment Scores:

Pred. NO.: 2.38e-30 Length: 81940
 Score: 810.50 Matches: 360
 Percent Similarity: 34.15% Conservative: 228
 Best Local Similarity: 20.91% Mismatches: 512
 Query Match: 11.76% Indels: 622
 DB: 24 Gaps: 54

US-09-818-990B-2 (1-1320) x ABK64829 (1-81940)

Qy 100 LeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGln 119
 Db 82 CTTAGTCCCAACCCCTTTTGGCCACTCTTGAGAAATCAGAGTG----- 123
 Qy 120 AspAsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGln 139
 Db 124 -----CCTAGAAAGATGACAACT-----CAAGCACCAGC 153
 Qy 140 TyrCysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGlu 159
 Db 154 TTTACGACGCGTTACAAAGCGTTGTGTA-----CTGGAG 189
 Qy 160 GluLeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLys 179
 Db 190 GGTAGTACCCGCAACCTTTGAGGCTCACATTAGTGTTCAGTTCTCGTAGGTGAGCTGG 249
 Qy 180 AsnHisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPhe 199
 Db 250 TTTAGGGATGCCAGGTGATTTC-----ACTTCCACTCTGCCCGGCGTCCACATCTCCTTT 306
 Qy 200 SerAspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThr 219
 Db 307 AGCGAT-----GGCCGCGCTAAACTGACGATCCCGCGGTGACT 345
 Qy 220 ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGlu 239
 Db 346 AAAGCCAACACAGTGGACGATATTCCTG----- 372
 Qy 240 GlnAlaAlaSerGluAlaAlaGlyLysAspThrThrProGlySerSerProSerSerLeu 259
 Db 373 -----AAAGCCCAACATGGATCTGGACAAGGCTAGTACTGTGAGCTTCTC 420
 Qy 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu 279
 Db 421 GTGAAAGCTGAGACA-----GCACCACCAACTTCGTTCAACGACTGCAGACATGACC 474
 Qy 280 ValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProProGln 299
 Db 475 GTGACAAAGAAAGCCCAAGTGAGACTCAAGTGAGACTGAGTGAATCCCTTAACCCCTGTG 534
 Qy 300 ValArgTrpTyrCysGluGlyLysLeuGluAsnSerProAspIleHisIleValGln 319
 Db 535 GTGAGTCTTACCGGGATGGACCGCAATCCAGAGCTCCCTTGATTTCCAAANTTCACAA 594
 Qy 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339

```
Db 595 GAAGGCGACCTCTACAGCTTACTGATTGCGAAGCATACCTGAGGACTCAGGACACCTAT 654
Qy 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
Db 655 TCAGTAAATGCCCAATAGCGTTGGAGAGCTACTTCGACTGCTGAATTAAGTTCTGTTCAA 714
Qy 360 Gly-----ValSerSerAspSerAspSerGluGlyAspProAsn 371
Db 715 GGTGAAGAAGTACCTGCTAAAGACAAAGACAATTTGTCGACTGCTCAGATCTCA 774
Qy 372 LysGluGluMetAsnArgIleGlnLys----- 380
Db 775 GAATCAAGACAAACCCGAATGAAAGAAAGATTGAAGCCCACTTTGATGCCAGATCAATT 834
Qy 381 -----Pro 381
Db 835 GCAACAGTTGAGATGGTCATAGATGGTGGCGCTGGGCAACAGCTGCCACATAAAACACCT 894
Qy 382 AsnGluValSerProProThrThrSerAlaValIleProProAlaValProGlnAla 401
Db 895 CCCAGGATTCTCCGAAGCCAAAGTCAAGATCCCAACACACCGCTGTATTGTGCGCCAAA 954
Qy 402 GlnHisLeuValAlaGlnProArgValAlaThrIleGlnCysGlnSerProThrAsn 421
Db 955 GCACAGCTGGCTCGGCAGCAGTCCCATCGCCATAAGACACACTCCCTTCCCGGTCAGA 1014
Qy 422 TyrLeuGlnGlyLeuAspGlyLysProIle----- 431
Db 1015 CAGGTGGGGCAGCGACCCCATCTCCGGTCAGGTCCGTCTCCAGCAGCAAGAATCTCC 1074
Qy 432 -----IleAlaAlaProValPheThrLysMetLeuGlnAsnLeuSer 445
Db 1075 ACATCCCCCATCAGGTCTGTGTAGTCTCCATTGCTCATGGTAAAGACTCAGGCATCCACC 1134
Qy 446 AlaSerGluGlnLeuValPheGluCysArgValLysGlyAlaProSerProLys 465
Db 1135 GTGCCACAGTCTCTGAAGTG-----CCTCCCCCT--- 1164
Qy 466 ValGluTrpTyrArgGluGly-----Thr 473
Db 1165 -----TGGAAAGCAAGAGGGCTAGTGGCCCTCTCATCTGAGGCTGAGATGAGAGAGACA 1218
Qy 474 LeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGlu 493
Db 1219 ACGTGCACACCTTACTCAGATCAGCAGCAGACAGAGATGGGAGGAGATACGGTGC 1278
Qy 494 ProGluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1279 CAGGCAAGTGACCATCAGTGGTGCTGCGGGTGTGCGGCCAGTGTGCGGCCAGTGCT 1338
Qy 514 ThrCysThrAla-----SerAsnLys 520
Db 1339 AGCTACACGACGACGAGGTGTGGCCACTGGTGTAAAGAGGTGAACAAAGATGCTGACAAA 1398
Qy 521 TyrGlyThrValSerIle----- 527
Db 1399 AGTCCAGCTGTGGCACTGTGTGTGCTGCCGTTCATATGCCACAGTGACAGAACCAAGTG 1458
Qy 528 -----AlaGlnLeuHisValArgGly 534
Db 1459 ATCAGCGCTGTAGACAGACTGCTCAGAGGACAACACGACTGCTGTGCACATCCAACT 1518
Qy 535 AsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAla 554
Db 1519 GCTCAAGAACAGGTGAAGAAAGGAAAGCGGAGAGAGACTGCTGTAACATAAGGTAGTAGTGC 1578
Qy 555 Ala-IleGluProGlnProSerProProHisSerGluPro----- 567
Db 1579 GCCGATAAAGCCAAAGAACAAAGAAATTAATCAAGAACCAAGAAATAATACCACAAG 1638
Qy 567 ----- 567
```

```
Db 1639 CAAGAGCAGATGCACGTAACTCATGAGCAGATAAGAAAAGAACTGAAAAAACATTTGTA 1698
Qy 568 -----ProSerValGluGlnProProLysProLys-----Le 578
Db 1699 CCAAGGTAGTAATTCGCGAGCTAAAG-----CCAAGAACAAAGAACTAGAAATTTCT 1752
Qy 578 uGluGlyValLeuValAsnHisAsnGlu----- 587
Db 1753 GAAGAAATTACTAGAAACACAAACAGTAACCTCAAGNAGCAATAANTCAAGGAAACTAGG 1812
Qy 588 -----ProArgSerSerSerArgIle-GlyL 596
Db 1813 AAAACAGTTGTACCTAAAGTCATAGTTGCCACACCCAAAGTCAAGAAAGTAAAGATTTAGTA 1872
Qy 596 eu-----ArgValHisPhe-----AsnLeuProGluA 605
Db 1873 TCAAGAGGTAGAGAGGCAATTACTTACCAAAAGAGAACAAAGTGCAAATAACTCAGGAGAAG 1932
Qy 605 spAspLysGlySer-Glu-----Ala 611
Db 1933 ATGAGAAGGAGGCGAGAAACTGCCTTGTCTACAATAGCAGTTGCTACTGCTAAAGCC 1992
Qy 612 SerSerGluAlaGlyValValThrArgGlnThr----- 623
Db 1993 AAGAACAAGAAACAATACTGAGAACTAGAGAAACTATGGCTACTAGACAAAGACAAATC 2052
Qy 623 ----- 623
Db 2053 CAAGTTACCATGGAAGGTGGACGTTGGAAAAAGGCTGAAGCTGTAGCAACAGTTGTT 2112
Qy 624 -----ArgProAspSer***GlnGluArg 631
Db 2113 GCTCCAGTAGACCAGGCCGAGTCAGAGAGCCAGAGAGCCTGGGCATCTTTGAGAAATCC 2172
Qy 632 PheAsnGlyGlnAlaThr-----LysThr 639
Db 2173 TATGCTCAGCAGACCACTTTGGAGTACGGATATAAGGAACGCATTTCCGCGCGCAAGGTA 2232
Qy 640 ProGlu---ProSerPheProValLysGluPro-----Pro 650
Db 2233 GCTCAGCTCTCCCAACGTCAGCCTCAGACCCCAACGCTGTCCCTAAAGCAGTCAAGCCT 2292
Qy 651 ProValLeuAlaLysProLys-----LeuAspSerThrGlnLeuGlnLeuHis 667
Db 2293 AGAGTAATCCAGGCTCTCTCTGAGACTATATCAAACTACTGATCAAAAGGAATGCAC 2352
Qy 668 AsnGlnValLeuLeuGluGln-----HisGlnLeu 677
Db 2353 ATATCATCACAGATCAAGAAAACTACAGATCTAACACGGAAGATTAGTCCATGTGGAT 2412
Qy 678 GlnAsnProProProSerSerProLysGluPheProPhe***MetThrValLeuAsn--- 696
Db 2413 AAACGCCCCGACAGCTAGCCCTCAC-----TTTACTGTTTCAAAAATTTCTGTT 2463
Qy 697 -----SerAsnAlaProAlaValThrThrSer*** 707
Db 2464 CCTAAGACAGAACATGGATATGAGGCATCAATAGCCGGTAGTGTATTTGCCACATTACAA 2523
Qy 708 LysGlnValLysAlaProSerSerGln-----ThrPheSerLeuAlaArgProLys 724
Db 2524 AAAGAGTTGTGAGCCACATCTTCTGCTCAGAAGATCACCAAAATCGGTGAAGGCTCTACT 2583
Qy 725 TyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSerSerSerProVal 744
Db 2584 GTG-----AAGCCCAAGTGAGACTAGAGTAAAGGCGAGAGCCACACCTTGCACAG 2634
Qy 745 PheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeu 764
Db 2635 TTCCCTTCGCTGACACACACAGACTACTACAG----- 2667
Qy 765 ValSerHisProSerValGlnThrLysSerProGlyLeuSerIleGlnAsnGluPro 784
Db 2668 --AGTGAAGCTGGCGTTGAGGTGAAAAAGGAAGTAGGGGTGAGCATCACTGGCACACC 2724
```

Qy 785 Leu----- 785
Db 2725 GTCCGTGAAGAGCGCTTTGAAGTACTGCAGCGACGCGAAGCAAGGTAAACAGAAACAGCA 2784
Qy 786 ---ProProGlyProThrGlu---ProThrProProProPheSer----- 800
Db 2785 AGAGTACCAGCAGCTGTTGAATTCCTCTTACTCCACCAACTTTGGTCTCGGCTTAAAA 2844
Qy 801 -----IleProSerGlyAsnGlnPheGlnProArgCys---ValSerProIlePro 816
Db 2845 AATGTGACTGTCATAGAAGGTGAATCTGTCACCTTGGAGTGCACATCTCTGGATACCACCA 2904
Qy 817 ValSerProThr-----SerArgIleGlnAsn 825
Db 2905 ---TCCCGCACAGTGCATGTGTACAGGGAAGAGTACCAAAATCGAAAGTTCCTCATGACTTC 2961
Qy 826 ProValAlaPheLeuSerSerVal----- 833
Db 2962 CAGATAACCTTCCAGAGTGGGAATTGCTCGTCTTATGATTCGGAAGCATTTTGGCGAAGAC 3021
Qy 833 ----- 833
Db 3022 ACCGGCGCATTTACTTGCAGTGTCTGTAATGAGGCTGGAACCGTCAGCACATCTCTGCTAT 3081
Qy 833 ----- 833
Db 3082 CTGGCTGTGCAGGTGTCAGAAAGATTTGAAAGGAACACACAGCGGTGACTGAGAAATTT 3141
Qy 834 -----LeuProSerLeuProAlaIleProProThrAsnAlaMet 846
Db 3142 ACTCAGAAAGAGAAACGCTTTGTGAGTCAAGAGATGTGGTTATGACTGTACTAGCTC 3201
Qy 847 ***LeuProArgSerAlaProSerMetProSer----- 857
Db 3202 ACAGAGGAACAGCAGGCGCTGGAGAACCTGCCGGCGCTTACTTTATTAACAAACCAAGTG 3261
Qy 858 ---GlnGlyLeuAlaLeuLysAsnThr-----LysSerPro 868
Db 3262 GTCCAGAAACTGGTGGAGGTGGGAGCGTGGTGTGGATGCCAAGTGGCGGCAACCCA 3321
Qy 869 GlnPro----- 870
Db 3322 AAGCCCATGTATACTGTGAAAAAATCTGGTGTCTCTTAACCACTGGATACAGATACAAA 3381
Qy 870 ----- 870
Db 3382 GTGAGTTACAAACAAACCGGTGAATGCAAGCTGGTGATTTCTATGACTTTTGTGCTGAT 3441
Qy 870 ----- 870
Db 3442 GATGCTGGAGAATACACTATTGTTGTCGCAATAAGCATGGAGAAACTTCTGCATCTGCT 3501
Qy 870 ----- 870
Db 3502 TCCTTGTGGAAGAGCTGATTATGAGTTACTGATGAAGTCCCAAGAGAAATGCTTTAT 3561
Qy 871 -----ValAsnAspAspAsnIleArgGluThr----- 879
Db 3562 CAGACAAAGTGAAGTGCATTTGTTCAAGAACCTGAAGTTGGAGAAACAGCACCTGGATT 3621
Qy 880 -----LysAsnAlaValIleArg---AspLeu 887
Db 3622 GTATACTCTGAGTATGAAAAAGAGTATGAAAAAGAACCAAGCCTTAATTTAGGAAGAAATG 3681
Qy 888 GlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGluThrLysIleSerSer 907
Db 3682 GCCAAAGATACTAGTGGTGCAGAACTTATGTAGAAGATCAGGAATTCATATTTCTTCC 3741
Qy 908 PheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrPro----- 924
Db 3742 TTTGAGAGAGACTTATTAAGAAATTTGAATATAGAATATAAAGACTACATATTAGAGAA 3801

Qy 925 ---ValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysIle 943
Db 3802 CTTCCTTGAAGAGATGGAGAAAGATGCGAGTGTGACATTTCTGAATCTGAAGCTGTT 3861
Qy 944 AlaProIlePheAspLysArgLysHisPheArgValThrGluGlySerProValThr 963
Db 3862 GAATCAGGATTTGATTTAAGAATCAAGAATATAGAATTTCTTGGGGGATGGGTGCTACT 3921
Qy 964 PheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGlyLys 983
Db 3922 TTTTCATGCAAGATGCTGGATATCCATTACCAAGATGCTTGGTACAAAGATGGCAAG 3981
Qy 984 GlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSer 1003
Db 3982 CGCATC---AAACATGGAGAAAGATACCAATGGACTTTTTCACAGATGGCAGAGTAGT 4038
Qy 1004 LeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsn 1023
Db 4039 CTGCGTATACCTGTTGTTCTTCCAGAAAGATGAAGGAATCTACACTGCATTTTGGCAGCAAT 4098
Qy 1024 ProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln----- 1037
Db 4099 ATTAAGGAATGCAATTTGCTCAGGGAAATTTGATGTGGAGCCTGCTGCACCACCTTGGGA 4158
Qy 1038 -----SerLeu----- 1039
Db 4159 GCTCCGACTTACATTCCACACTAGAGCCAGTACAGCAATCAGATCTCTCTCCACGT 4218
Qy 1040 -----ProIleArg-----SerArgLeuThr 1046
Db 4219 TCAGTGAGCAGGTCTCCTATACGCATGCTCTCTGCACGGATGCACCTGCAAGGATGCT 4278
Qy 1047 SerAlaGlyGlnSerHisArgGlyArgSer-----ArgValGlnGluArgAspLys 1063
Db 4279 CCTGCAGGATGCTCCCTGCAAGAAATGTCCCTCGACGTAGGCTGGAGGAGACAGATGAG 4338
Qy 1064 GluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetVal 1083
Db 4339 TCACAACCT---GAGAGACTATATAAACCCAGCTTTGTGTTTAAAAACCTGTTCTTTCAA 4395
Qy 1084 AlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGlu 1103
Db 4396 TGTTTAAAGGGGCAAACTGCAGATTTGACTTAAAGGTGTGTGTAGACCTTATGCCAGAG 4455
Qy 1104 LeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuVal 1123
Db 4456 ACGTCTGCTGTTTCATGATGGCCACCAATGTCAATGACTATACCCATAAAGTAGTCATT 4515
Qy 1124 ArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThr 1143
Db 4516 AAAGAAGATGCTACTCAATCACTAATATTGTCCTGCCACACCCAGTGAATCTCTGGGAA 4575
Qy 1144 TyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuLeuLeuSerVal 1163
Db 4576 TGGACTGTGGTGGCCAAAACAGGGCAGAGACTCTCAATTTTCAGTGAATTTTAACTGTG 4635
Qy 1164 ValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyVal 1183
Db 4636 GAAGCTGTGGAAACATCAGGTAAACCCGATGTTGTAGAAAACTGAAAAATGTCAATATA 4695
Qy 1184 ProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProValPhe 1203
Db 4696 AAGGAAGTTCCTCCGACTTGAATGAAATGAAATGACAGCTACGGGTAAACCCCAACCTGACATT 4755
Qy 1204 TyrTrpLysLysAspAsnGluThrIle---ProCysThrArgGluArgIleSerMetHis 1222
Db 4756 GTATGGTTGAAAAACAGTGCATCATTTGCTGCCTCATAAATATCCAAATCAGAAATT--- 4812
Qy 1223 GlnAspThrThrGlyTyrAlaCysLeuIleGlnProAlaLysLysSerSerAlaGly 1242
Db 4813 GAAGGAACCAAGGAGAGAGCTGCCCTTAAATCCATCCACTGTCCAGCAAGATTCGCC 4872
Qy 1243 TrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 1262

```
||||| 4873 TGTATACGCGACTGCTATTATAAAGCTGGCAGACACACTACAAGATGCAAGTAAAT 4932
||||| 1263 IletyAlaGlntrpHisGlnIleProProMetSerValArgProSerGlySer 1282
||||| 4933 GTTGAAGTTGAGTTGTCAGAGCCTGAGCCAGAGAGAAAGTTAATCATCTCCACCGGGGACA 4992

RESULT 11
AAS05390
ID AAS05390 standard; DNA; 81940 BP.
XX AAS05390;
XX 24-OCT-2001 (first entry)
DE Human titin (connectin) gene sequence.
XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW titin-related disease; zebrafish; heart failure; heart disease; ds.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 133..80913
FT /*tag= a
FT /product= "titin"
XX
XX WO200151666-A1.
XX 19-JUL-2001.
XX 12-JAN-2001; 2001WO-US01212.
XX 12-JAN-2000; 2000US-0175787.
XX (GEO ) GEN HOSPITAL CORP.
XX Fishman MC;
XX WPI: 2001-451869/48.
XX P-PSDB; AAU05396.
XX
XX Determining if a subject has or is at risk of developing a
XX titin-related disease or condition, particularly heart failures,
XX comprises detecting the presence of a mutation in the titin gene
XX
XX Disclosure; Page 35-57; 114pp; English.
XX
XX The present sequence encoding for human titin (also known as connectin)
XX is described in an invention relating to a novel method for determining
XX whether a subject has or is at risk of developing a titin-related
XX disease or condition. The method comprises analysing a nucleic acid of
XX a sample from the subject and detecting the presence of a mutation
XX (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
XX titin gene, which indicates that the subject has or is at risk of
XX developing a titin-related disease. The zebrafish which has a phenotype
XX similar to mammalian heart failure is used as a model. The method is
XX useful for detecting an increased likelihood of heart disease, such as
XX heart failure, in a patient, so that appropriate intervention can be
XX instituted before any symptoms occur. The method may also be used to
XX facilitate determination of etiology of an existing heart condition,
XX such as heart failure, to identify compounds that can be used to treat
XX or prevent heart conditions, in prenatal genetic screening, e.g. to
XX identify parents who may be carriers of a recessive titin mutation.
XX Compounds identified using the methods may be used to treat patients
XX that have or are at risk of developing heart disease, e.g. heart
XX failure.
XX
XX Sequence 81940 BP; 26373 A; 17100 C; 18899 G; 19568 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3,78e-30 Length: 81940
XX Score: 806.50 Matches: 359
```

```
Percent Similarity: 34.15% Conservative: 229
Best Local Similarity: 20.85% Mismatches: 512
Query Match: 11.71% Indels: 622
DB: 22 Gaps: 54

US-09-818-990B-2 (1-1320) x AAS05390 (1-81940)

QY 100 LeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGln 119
||||| 82 CTTAGTCCCACCTTTTAGGCACTCTTGAGAAATCAGAGTG----- 123
QY 120 AspAsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGln 139
||||| 124 -----CCTAGAAAGATGACAAC-----CAAGCACCGACG 153
QY 140 TyrCysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAppHeileGlu 159
||||| 154 TTTACGACGCGTTACAAAGCGTTGTGTA-----CTGGAG 189
QY 160 GluLeuSerSerLysHisSerHisSerLysArgIleArgProArgAlaCysLys 179
||||| 190 GGTAGTACCGCAACCTTTGAGGCTCATTAGTGGTTTCCAGTCTCTGAGTGAGCTGG 249
QY 180 AsnHisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPhe 199
||||| 250 TTTAGGATGCCAGGTGATTTC-----ACTTCACCTCTGCCGCGTCGAGATCTCCTTT 306
QY 200 SerAspLeuSerGluArgArgGluArgSerSerValProIleProAlaAspThr 219
||||| 307 AGCGAT-----GGCCGCGCTAAACTGACGATCCCGCGGTGACT 345
QY 220 ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnAlaLysArgGluAlaGlu 239
||||| 346 AAAGCCCAACAGTGGACGATATTCCTG----- 372
QY 240 GlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerLeu 259
||||| 373 -----AAAGCCCAACATGGATCTGGACAACGACTAGTACTGCTGAGCTTCTC 420
QY 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu 279
||||| 421 GTGAAAGCTGAGACA-----GCACCACCACTTCCTTCAACGACTGCAGAGATGACC 474
QY 280 ValProGluGlyThrArgValGlnLeuAspCysIleValGlyIleProProGln 299
||||| 475 GTGAGACAAGCAAGCAAGTGAGACTCCAAGTGAGAGTGACTGGAATCCCTACACCTGTG 534
QY 300 ValArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln 319
||||| 535 GTGAAGTCTACCGGGATGGAGCGAAATCCAGAGCTCCCTTGATTTCCAAATTTCAAA 594
QY 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339
||||| 595 GAAGGACCACTTACAGCTTACTGATTGCAAGACATACCTGAGGACTCAGGGACCTAT 654
QY 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
||||| 655 TCAGTAATGCCACCAATACGTTGGAGAGAGCTACTTTCGACTGCTGAATTTACTGGTTCAA 714
QY 360 Gly-----ValSerSerAspSerGluGlyAspProAsn 371
||||| 715 GGTGAAGAAGATGACCTGCTTAAAGACAAAGACAAATTTGTCGACTGCTCAGATCTCA 774
QY 372 LysGluGluMetAsnArgIleGlnLys----- 380
||||| 775 GAATCAAGACAAACCCGATTTGAAAGAGAGATTGAGCCCACTTTGTGTCAGATCAATT 834
QY 381 -----Pro 381
||||| 835 GCAACAGTTGAGATGGTCATAGATGGTCCGCTGGGCAACAGCTGCCACATATAAACACCT 894
QY 382 AsnGluValSerProProThrThrSerAlaValIleProProAlaValProGlnAla 401
||||| 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
```

Db 895 CCCAGGATTCTCCGAAGCCMAAGTCAGATCCCCACACACCACCGTCTATTGCTGCCAAA 954
Qy 402 GlnHisLeuValaLagInProArgValaLathrIleGlnGlnCysGlnSerProThrAsn 421
Db 955 GCACAGCTGGCTCGCAGCAGCTCCCATAGACACTCCCTTCCCGGTGAGA 1014
Qy 422 TyrLeuGlnGlyLeuAspGlyLysProIle----- 431
Db 1015 CACGTGGGGCAGCAGCCCATCTCCCGTCAAGTCCGTGCTCCAGCAGCAAGAATCTCC 1074
Qy 432 -----IleAlaAlaProValPheThrLysMetLeuGlnAsnLeuSer 445
Db 1075 ACATCCCCATCAGGTCTGTAGCTTCATGTCTATGCGTAAGACTCAGGCATCCACC 1134
Qy 446 AlaSerGluGlyGlnLeuValValPheGluCysArgValLysGlyAlaProSerProLys 465
Db 1135 GTGCCACAGCTCTGAGTG-----CCTCCCCCT--- 1164
Qy 466 ValGluTrpTyrArgGluGly-----Thr 473
Db 1165 -----TGAAGCAAGAGGCTAGTGGCTCTCATCTGAGGCTGAGATGAGAGACA 1218
Qy 474 LeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGlu 493
Db 1219 ACGTGTGAACCTCTACTCAGATCAGACAGACAGAGAGATGGGAAGGAGATACGGGTGC 1278
Qy 494 ProGluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1279 CAGGAGCAAGTACCATCAGTGGTCTGCGGGTCTGCCCGCAGTGTGCGGCAGTGCT 1338
Qy 514 ThrCysThrAla-----SerAsnLys 520
Db 1339 AGCTACGCAGCAGAGCGTGTGCCACTGGTCTAAAGAGGTGAACAAGATGCTGCACAAA 1398
Qy 521 TyrGlyThrValSerSerIle----- 527
Db 1399 AGTCGAGTGTGGGACTGTTGTGCGCTTGATATGCCAGAGTGAGAGAACCAAGTG 1458
Qy 528 -----AlaGlnLeuHisValArgGly 534
Db 1459 ATCAGCGCTGTAGACAGACTGCTCAGAGGACACACGACTGCTGTGCGCATCCACCT 1518
Qy 535 AsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAla 554
Db 1519 GCTCAAGAACAGGTAAAGAAAGGAAGCGAGAAAGACTGCTGTAAGTAGTAGTGCC 1578
Qy 555 Ala-IleGluProGlnProSerProProHisSerGluPro----- 567
Db 1579 GCCGTAAGCCAGGAACAGAAATTAATTAAGAACCAAGAAATAATTACCACAAAG 1638
Qy 567 ----- 567
Db 1639 CAAGAGCAGATGCACGTAACATCATGACAGATAAGAAAGAAAGAAATTTGTA 1698
Qy 568 -----ProSerValGluGlnProProLysProLys-----Le 578
Db 1699 CCAAAGGTAGTAATTTCCGCGAGCTAAAG-----CCAAAGAAACAAGAAACTAGAAATTTCT 1752
Qy 578 uGluGlyValLeuValAsnHisAsnGlu----- 587
Db 1753 GAAGAAATTAAGAAACAAGAACTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1812
Qy 588 -----ProArgSerSerArgIle-GlyL 596
Db 1813 AAAACAGTTGTACCTAAAGTCATAGTTGCCACACCCCAAGTCAAGAAAGAAAGTAATTTAGTA 1872
Qy 596 eu-----ArgValHisPhe-----AsnLeuProGlu 605
Db 1873 TCAAGAGGTAGAGAGCAATTACTACCAAAAGAGAACAAAGTGCATAACTCAGGAGAAG 1932
Qy 605 spAspLysGlySer-Glu-----Ala 611
Db 1933 ATGAGAANGGAAGCCGAGAAACTGTCCTGTACAAATAGCAGTTGCTACTGCTAAAGCC 1992

Qy 612 SerSerGluAlaGlyValValThrThrArgGlnThr----- 623
Db 1993 AAAGAAACAAGAAACAATACTAGAGAACTATGCTACTAGACAAGAAACAATC 2052
Qy 623 ----- 623
Db 2053 CAAGTTACCATGAAAGGTGGAGCTTGGAAAAAGGCTGAAGCTGTAGCAACAGTTGTT 2112
Qy 624 -----ArgProAspSer***GlnGluArg 631
Db 2113 GCTGCAGTAGACAGGCCCGCAGTCAGAGAGCCAGAGAGCTTGGGCATCTTGAAGAATCC 2172
Qy 632 PheAsnGlyGlnAlaThr-----LysThr 639
Db 2173 TATGCTCAGCAGACCACTTTGGAGTACGGATATTAAGAACGCATCTTCCGCCGCAAGSTA 2232
Qy 640 ProGlu---ProSerPheProValLysGluPro-----Pro 650
Db 2233 GCTGAGCCTCCCAACGTCAGGCTCAGAACCCACGCTGTCCCTAAAGCAGTCAAGCCT 2292
Qy 651 ProValLeuAlaLysProLys-----LeuAspSerThrGlnLeuGlnGlnHis 667
Db 2293 AGAGTAATCCAGGCTCCTCTGTGAGACTCATATAAACTACTGATCAAAAGGGGAATGCAC 2352
Qy 668 AsnGlnValLeuLeuGluGln-----HisGlnLeu 677
Db 2353 ATATCATCAGATCAAGAAACTACAGATCTAACACGGAAAGATTAGTCCATGTGGAT 2412
Qy 678 GlnAsnProProProSerSerProLysGluPheProPhe***MetThrValLeuAsn--- 696
Db 2413 AAAGCCCGCGCAGCTAGCCCTCAC-----TTTACTGTTCAAAAATTTCTGTT 2463
Qy 697 -----SerAsnAlaProAlaValThrThrSer*** 707
Db 2464 CTTAAGACAGAACATGATATGAGGCATCAATAGCCGCTAGTGTATTTGCCATTTACAA 2523
Qy 708 LysGlnValLysAlaProSerSerGln-----ThrPheSerLeuAlaArgProLys 724
Db 2524 AAAGAGTGTGACCCACATCTTCTGCTCAGAGATCACCAATCGGTGAGGCTCCTACT 2583
Qy 725 TyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSerSerProVal 744
Db 2584 GTG-----AAGCCAGTAGACTAGAGTAGAGGCGAGAGCCACACCTTTGCCACAG 2634
Qy 745 PheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeu 764
Db 2635 TTCCCTTCGCTGCACACACAGATACCTTACAAG----- 2667
Qy 765 ValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluPro 784
Db 2668 ---AGTCAAGCTGGCGTTGAGTGAAGGAAAGTAGGAGGTGAGCATCCTGGCACCACC 2724
Qy 785 Leu----- 785
Db 2725 TCCCGTAGAGAGCGCTTTGAAGTACTGCAGCGAGCGAAGCCAGAGTAACAGAAACAGCA 2784
Qy 786 ---ProProGlyProThrGlu---ProThrProProPheThrPheSer----- 800
Db 2785 AGAGTACACGACCTGTTGAATTCCTGTACTCCACCAACTTTGGTCTCGGGCTTAAAA 2844
Qy 801 -----IleProSerGlyAsnGlnPheGlnProArgCys---ValSerProIlePro 816
Db 2845 AATGTGACTGTCATAGAGGTGAATCTGTCCACCTTGGAGTGGCCACATCTCTGGNATACCA 2904
Qy 817 ValSerProThr-----SerArgIleGlnAsn 825
Db 2905 ---TCCCGACAGTGACATGGTACAGGGAAGACTTACCAAAATCGAAAGTTCCATTGACTTC 2961
Qy 826 ProValAlaPheLeuSerSerVal----- 833
Db 2962 CAGATAACCTTCCAGAGTGAATTCGCTTATGATTTCGGAAGCAATTTGCGGAAGAC 3021

```
QY 833 ----- 833
Db 3022 AGCGGCGATTACTTGCAGTGTGTAATGAGCGTGAACCGTFCAGACATCTGTCTAT 3081
QY 833 ----- 833
Db 3082 CTGGCTGTGCAGGTGTCAGAAGATTTTGAAAAGGAAACACACAGCCGTGACTGAGAAATTT 3141
QY 834 -----
Db 3142 ACTACAGAAGAAACCGCTTTGTGAGTCAAGAGATGTTGATGACTGATAGTGCCTC 3201
QY 847 ***LeuProArgSerAlaProSerMetProSer----- 857
Db 3202 ACAGAGAAACAGCAGCGCTGGAGAACCTGCCCGCCTTACTTTATTACAAAACACAGTG 3261
QY 858 ---GlnGlyLeuAlaLysLysAsnThr-----LysSerPro 868
Db 3262 GTCAGAAACTGGTGAAGGTGGGAGCGTGGTGTGGATGCCAAGTTGGCGGCAACCCA 3321
QY 869 GlnPro----- 870
Db 3322 AAGCCCCATGTATACTGGAATAATCTGGTGTCTCTAACCACTGGATACAGATACAAA 3391
QY 870 ----- 870
Db 3382 GTGAGTTACAACAACAACCGGTGAATGCAAGCTGGTGTCTCTATGACTTTTGTCTGAT 3441
QY 870 ----- 870
Db 3442 GATGCTGGAGAATACACTATTGTTGTCGAATAAGCATGAGAAACTTCTGCATCTGCT 3501
QY 870 ----- 870
Db 3502 TCCTGTGTTGAAGAGCTGATTATGAGTTACTGATGAAGTCCCGAGCAAGAAATCCTTTAT 3561
QY 871 ----- 879
Db 3562 CAGACACAAGTGACTGCATTGTTCAAGAACCTGAAGTTGGAGAAACAGCACCTGGATT 3621
QY 880 -----LysAsnAlaValIleArg---AspLeu 887
Db 3622 GTATFACTGTGATGATAAAGAGATGATAAAGAAAGCAAGCCTTAATTAGGAAGAAATG 3681
QY 888 GlyLysIleThrPheSerAspValArgProAsnGlnGluTyrLysIleSer 907
Db 3682 GCCAAAGACTGTAGTGGTCAGAACTTGTAGAAATCAGGAATTCATATTTCTTCC 3741
QY 908 PheGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrPro----- 924
Db 3742 TTTCAAGAGAGACTTATTAAGAAATTCATATAGATAATAATAAGACTACATTAGAAGAA 3801
QY 925 ---ValAspGluSerAspAspGluIleGlnHisaspGluIleProThrGlyLysCysIle 943
Db 3802 CTTCTTGAAGAAGATGGAGAGAAAGATGGCAGTTGACATTTCTGAATCTGAAGCTGT 3861
QY 944 AlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValThr 963
Db 3862 GAATCAGGATTTGATTAAGAAATCAAGAAATATAGAAATTTGAGGGGATGGGTGCTACT 3921
QY 964 PheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGlyLys 983
Db 3922 TTTTCATTGCAAGATGCTGGATATCCATTACCAAGATGCTTGTGTACAAAGATGGCAAG 3981
QY 984 GlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSer 1003
Db 3982 CGCATC---AAACATGGAGAAAGATACCAATGGACTTTCTACAAGATGGCAGAGCTAGT 4038
QY 1004 LeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsn 1023
Db 4039 CTGGTATACCTGTGTCTTCCAGAAGATGAAGAAATCTACACTGCATTTGCCAGCAAT 4098
QY 1024 ProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln----- 1037
```

```
Db 4099 ATTAAGGAAATGCAATTTGCTCAGGAAATTTGATGCGAGCGCTGCTGCACCACCTTGA 4158
QY 1038 -----SerLeu----- 1039
Db 4159 GCTCCGACTTACATTCACACACTAGACCCAGTGAAGCAATCAGATCTCTCTCTCCAGT 4218
QY 1040 -----ProIleArg-----SerArgLeuThr 1046
Db 4219 TCAGTGAGCAGGTCTCTATACGATGTCTCTCCAGCGATGTCCACCTGCAAGGATGTCT 4278
QY 1047 SerAlaGlyGlnSerHisArgGlyArgSer-----ArgValGlnGluArgAspLys 1063
Db 4279 CCTGCAGGATGTCCCTGCAAGAAATGTCCTCCAGCTAGCTGGAGGACACAGATGAG 4338
QY 1064 GluProLeuGlnArgPhePheArgProHisPheLeuGlnAlaProGlyAspMetVal 1083
Db 4339 TCACAACTT---GAGAGACTATATAAACCACTTTGTTGTTAAACCTGTTTCTTCAAA 4395
QY 1084 AlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGlu 1103
Db 4396 TGTTTGAAGGGCAAACTGCCAGATTTGACTTAAAGGTGTTGGTAGACCTATGCCAGAG 4455
QY 1104 LeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuVal 1123
Db 4456 ACGTCTGTTGTTGATGCGCAGCAAAATGTCAATGACTATATCCCATAAAGTAGTCAT 4515
QY 1124 ArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThr 1143
Db 4516 AAAGAAGATGGTACTCACTCAATTAATTTGTCCTGCCACCCAGCTGATTTCTGGGAA 4575
QY 1144 TyrLysCysIleAlaThrAspLysThrGlyGlnAsnSerPheSerLeuGluLeuSerVal 1163
Db 4576 TGGACTGTGGTGGCCAAACACAGCGGAGGAGATCTCAATTTTCAGTGATTTTAACTGTG 4635
QY 1164 ValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyVal 1183
Db 4636 GAAGCTGTGGAACATCAGGTAAAACCGATGTTGTAGAAAAAATGMAAAATGTCAATATA 4695
QY 1184 ProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProProValPhe 1203
Db 4696 AAGGAAGGTTCCTCCAGCTGGAAGTCAAGCTACGGGTAAACCCCAACCCAGCTGACAT 4755
QY 1204 TyrTrpLysLysAspAsnGluThrIle---ProCysThrArgGluArgIleSerMetHis 1222
Db 4756 GTATGGTTGAAAACAGTGACATCATGTGCTCTATAAATATCCAAAATCAGAAAT--- 4812
QY 1223 GlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGly 1242
Db 4813 GAAGGAACCAAGGAGAGAGCTGCCCTTAAATTCGATTCCTCACTGTCCAGCAAGATTCGCC 4872
QY 1243 TrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 1262
Db 4873 TGGTATATCTGGCAGTGTATTAATAAAGCTGGCAGACACATACAAGATGCAAGATTAAT 4932
QY 1263 IleTyrAlaGlnTyrHisGlnIleProProMetSerValArgProSerGlySer 1282
Db 4933 GTTGAAGTTGATTTGCGAGACCTTGAGCCAGAGAAAGTTAATCATCCACGGGGGACA 4992
RESULT 12
AXA40026
ID AXA40026 standard; DNA; 790 BP.
XX
AC AXA40026;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
```


PS Disclosure; Page 23-25; 43pp; English.

XX The present sequence encodes a human myotilin (myofibrillar protein
CC with titin-like Ig-domains) protein. The myotilin protein is a component
CC of the striated and cardiac muscle cytoskeleton. The protein contains two
CC C2-type immunoglobulin (Ig)-like domains with homology to certain
CC it localises within the I-bands and is bound to alpha-actinin, and along
CC the sarcolemmal membrane. The myotilin gene locates in chromosome 5q31
CC inside a 2 Mb region, which contains the LGMD1A disease gene.
CC Transfection of myotilin into mammalian cells induces formation of
CC thick actin bundles. The myotilin protein is used for preparing compounds
CC that regulate actin-cytoskeleton and cell growth, and cancer and
CC microbial infections. The protein is also used in the preparation of
CC medicines for treating cancer and antimicrobial infection.

XX Sequence 2244 BP; 737 A; 506 C; 416 G; 585 T; 0 other;

Alignment Scores:

Pred. No.: 1.62e-29 Length: 2244
Score: 760.00 Matches: 220
Percent Similarity: 45.95% Conservative: 98
Best Local Similarity: 31.79% Mismatches: 220
Query Match: 11.03% Indels: 155
DB: 21 Gaps: 19

US-09-818-990B-2 (1-1320) x AAZ61262 (1-2244)

QY 653 LeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeu 672
DB 98 CTACCAAGCGCAGGAGCAGCAGTAGTATCTCAGGATCTCAACAAGGAGAGCAGCAAGGTT 157
QY 673 GluGlnHisGlnLeuGlnAsnProProSerSerProLysGluPheProPhe**Met 692
DB 158 GCTTCTGATCTCTACAA-----CCTCGGTAAATCCAGGCTGTGGCC----- 202
QY 693 ThrValLeuAsnSerAsnAlaProAlaValThrThrSer**LysGlnValLysAla 712
DB 203 -----CAAAATTCAGGCGCCACCCTCCAGGAACAATCATATAGTAATAATTG 253
QY 713 ProSerSerGlnThrPheSerLeuAla-----ArgProLysTyrPhePhe 727
DB 254 CCTTCATCTTCCATATACCACTAAGCATGTTAACTACGAACTGCCAAACACTTCATC 313
QY 728 ProSerThrAsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeu 747
DB 314 CAGTCCAAACCCATGCTGGCTCGAGATTCGAGCTCCTCGACCAGAACTCCAGCTTC 373
QY 748 SerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuVal-SerHi 767
DB 374 TCTAGC-----CAGACCAACAG-----TCTTCCATATATCATCCAGCCC 412
QY 767 sProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProPr 787
DB 413 CGCCAGGTACAGAGCAAA-----GATTTCCTGCCTCC 445
QY 787 o-----GlyProThrGluProThrProProProPheThrPheSerIleProSerGl 804
DB 446 TCACACTGAGCTCTCAATACATCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 505
QY 804 YAsn-GlnPheGlnProArgCysValSerProIleProValSer-ProThrSerArgIle 823
DB 506 CATGCTGGCTCCACCCAGG-----CCAAAGGTTACAAACCACTATATAC 550
QY 824 GlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIleProProThr 843
DB 551 CAGTCCCGCAGCGACTCTCCAGCTCAGCTCATATATACATACAG-----CCTGATTAC 601
QY 844 AsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLys 863
DB 602 AATGACAGTAAATTCCTCGCTATGATTCGAATTCGAATTCGAATTCGAATTCGAATTCGA 661
QY 864 -----AsnThrLysSerProGlnProValAsn-----Asp 873

DB 662 CCTATAAATGCAAGCCATCCCAAAATGCAAAATGCTAAGCCCATACCAAGAACTCCTGAT 721
QY 874 ASPAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPhe 893
DB 722 CATGAATAACAGGATCAAAAGAGAGCTTTGATTCAAGATTTGGAAGAAAGCTGAAATGC 781
QY 894 SerAspValArgProAsnGlnGlnGluTyrLysIleSerPheGluGlnArgLeuMet 913
DB 782 AAGGACACC-----CTTCTTCAATATGGAATCAACCTCTCAACA 820
QY 914 AsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGln 933
DB 821 TATGAAGA-GAA----- 831
QY 934 HisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHis 953
DB 831 ----- 831
QY 954 PheArgValThrGluGlySerProValThrPheThrCysLysIleValGlyIleProVal 973
DB 831 ----- 831
QY 974 ProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
DB 832 -----GATGGCTCGCAGATTCTAGGACCACAGAA----- 861
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSer-AspAs 1013
DB 862 -----TGCAGCTGCTGTGTTTCAAGC-----TCAGGATGA 891
QY 1013 pAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHi 1033
DB 892 CAGTGGT-----GCACAAGACTCCGACCAACACCACTCAGAACATGCGCG 936
QY 1033 sLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisAr 1053
DB 937 ACTGCAAGTTCTTACATACACAAAGTAAGAGTAGATCAACCTCAAGGGAGATGTGAAT-- 994
QY 1053 gGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPr 1073
DB 995 -----GATCAGGATGCAATCCAGAGAAATTTTCCACC 1029
QY 1073 oHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAs 1093
DB 1030 ACGTTTCATTCAAGTGCAGAGAACATGTCGATTGATGAGGAAGATCTCGACAANTGGA 1089
QY 1093 pCysLysValSerGlyLeuProProProGluLeuThrTrpLeuLeuAsnGlyGlnProVa 1113
DB 1090 CTTCAAGTGCAGTGCAGCTCCAGCTCCTGATGTCATGTTCTATAAATGGAAGAACAGT 1149
QY 1113 lLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIl 1133
DB 1150 TCAATCAGATGATTGCACAAAATGATAGTGTCTGCAAGGGTCTTCATTCACCTCATCTT 1209
QY 1133 eAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGl 1153
DB 1210 TGAAGTAGTCAGACTTCAGATCGAGGGCTTATGATGTTGCCAAGATATAGAGCAGG 1269
QY 1153 yGlnAsnSerPheSerLeuGluLeuSerValAlaLysGluValLysLysAlaProVa 1173
DB 1270 AGAAGCCACCTTCACTGTGCAAGTGGATGCTCTGCAAAAGAACATATAAGAGCACCAT 1329
QY 1173 lLeuLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCy 1193
DB 1330 GTTTATCTACAAACACACAGAGCAAAAGTTTATAGGGAGATTCAGTGAAACTAGATG 1389
QY 1193 sArgValIleGlyMetProProProValPheTyrTrpLysLysAspAsnGluThrIlePr 1213
DB 1390 CCAGATCTCGGCTATACCTCCCAACAGCTTCTCGAAGAAAGATATAGAAATGGGTACA 1449
QY 1213 oCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIl 1233

Db	1450	ATTCAACACTGACCGAATAAGCTTATATCAAGATAACACTGGGAAGAGCTTACTTTACTGAT	1509
Qy	1233	eGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaG	1253
Db	1510	AAAAGATGTAACAAGAAAGATGCTGGTGGTATACTGTGTCAGCAGCTTAATCAAGCTGG	1569
Qy	1253	yIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIleProPr	1273
Db	1570	AGTGACTACATGTAACACAAAGATTAGACGTTACGGCAGCTCCAAACCAAACTCTCCAGC	1629
Qy	1273	oPro-----MetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysG	1291
Db	1630	TCCTAAGCAGCTTACGGGTTCCGACCACATTCAGCAAAATATTTAGCAGCTTAATGGGAAAG	1689
Qy	1291	yLeuAspIlePheSerAlaPheSerSerMetGluSerThrMetValTyrSerCysSerSe	1311
Db	1690	TTTGAATGTAACAACAGCTTTTAAACCCA--GAAGGAGAAATTTTCAGCGTTTGCAGCTCA	1746
Qy	1311	rArgSerValValGluSerAspGluLeu	1320
Db	1747	ATCTGGAGCTCTATGAAGAGTGAAGAACTT	1774
RESULT	14		
AXX	AAAX40027/c		
ID	AXX40027	standard; DNA; 747 BP.	
XX	XX		
AC	AXX40027;		
DT	XX		
DT	02-JUL-1999	(first entry)	
DE	XX	Prostate cancer associated gene.	
DE	XX		
XX	XX	Cancer associated antigen; diagnosis; research; treatment; human;	
KW	KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW	KW	prostate cancer; ss.	
XX	XX		
OS	XX	Homo sapiens.	
PN	WO9904265-A2.		
XX	XX		
PD	28-JAN-1999.		
XX	XX		
PF	15-JUL-1998;	98WO-US14679.	
PR	22-JUN-1998;	98US-0102322.	
PR	17-JUL-1997;	97US-0896164.	
PR	10-OCT-1997;	97US-0061599.	
PR	10-OCT-1997;	97US-0061765.	
PR	10-OCT-1997;	97US-0948705.	
PR	11-OCT-1997;	97GB-0021697.	
XX	(LUDW-) LUDWIG INST CANCER RES.		
XX	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;		
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;		
PI	Tureci O;		
XX	WPI; 1999-132448/11.		
XX			
DR	New isolated cancer associated nucleic acids and polypeptides -		
PT	isolated using sera from cancer patients, used to develop products		
PT	for the diagnosis, monitoring or treatment of cancers		

The diagnosis, monitoring, research, or treatment of conditions characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer.

XX

SQ Sequence 747 BP; 156 A; 180 C; 192 G; 218 T; 1 other;

Alignment Scores:

Pred. No.:	1_93e-25	Length:	747
Score:	668.50	Matches:	130
Percent Similarity:	75.81%	Conservative:	33
Best Local Similarity:	60.47%	Mismatches:	48
Query Match:	9.70%	Indels:	4
DB:	20	Gaps:	2

US-09-818-990B-2 (1-1320) x AAX40027 (1-747)

QY	1109	ASnGlyInProValLeuProAspAlaSerHisLysMetLeuValArg-GluThrGlyVa	1128
Db	745	GATGGAAAGCGGTACGCCCTTGACAGGGCTCCCAAGATGCTGGTGAAGAAGCGGGT	686
QY	1128	LHisSerLeuLeuIleAspProLeuThrGlnArgspAlaGlyThrTyfLysCysileal	1148
Db	685	GCATCTCTCATAGAGCACGTCACGTCAGTGTGCGGCATCTACACATGTATAGC	626
QY	1148	aThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValvalAlaIalySgluVa	1168
Db	625	TACCACCGAGCAGGACAGAACATCATTCAGCCTGGAGCTGTGGTTGCTGAAGAAGC	566
QY	1168	LlysLysAlaproValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisPr	1188
Db	565	ACACAACCCCCTGTGTTATTGAGAAGCTCCAACACAGAGTTCGTGATGGGTACCC	506
QY	1188	oValArgLeuGluCysArgValIleGlyMetProProProValPheTyTrpLysLysas	1208
Db	505	AGTGGCGCTGGAATGCTGTATTGGGAGTGCCACCACCTCAGATATTTTGAAGAAGA	446
QY	1208	pAsnGluThrIleProCystThrArgGluArgIleSerMetHisGlnAspThrThrGlyTy	1228
Db	445	AAATGAATCACTCACTCACGACTGACCGAGTGAGCATGCCACGAGCAACACCGGTA	386
QY	1228	rAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyTrpLeuSerAl	1248
Db	385	CATCTGCCTGCTCATTCAGGAGCCACAAAAGAAGATGCTGGGTGCTACTGTGTACG	326
QY	1248	aLysAsnGluAlaGlyIleValSerCystThrAlaArgLeuAspileTyfAlaGlnTrphI	1268
Db	325	CAAGAATGAAGCAGGGATTGTCTCTACTGTCAGGCTGGACGTTTACACCCAGTGGA	266
QY	1268	sHis-----GlnIleProProMetSerValArgProSerGlySerArgTyfGlySe	1286
Db	265	TCAGCAGTCACAGCAGCACCAAGCAAAAAAAGATACGGCCCTCAGCCAGTCGTATG	206
QY	1286	rLeuThrSerLysGlyLeuAspilePheSerAlaPheSerSerMetGluSerThrMetVa	1306
Db	205	ACTTTCGACACGAGGACTGACATCAAAAGCAGCGTTCCAACT--GAGGCCAACCCATC	149
QY	1306	lTySerCysSerSerArgSerValValGluSerAspGluLeu	1320
Db	148	TCACCTGACACTGAATACTGCCTTGGTAGAAGATGAGGACCTG	106

RESULT 15
AAC75084
ID AAC75084 standard; cDNA; 392 BP.
XX AAC75084;
XX AC
XX AC
DT 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF639 polynucleotide sequence SEQ ID NO:1277.
XX

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatologic; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antinaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

XX PN WO200058473-A2.

XX
PD 05-OCT-2000.

XX .
PF 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999: 99US-0127607.

PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728

PR 30-MAR-2000; 2000US-0540763.
XX

PA (CURA-) CURAGEN CORP.
YY

PI Shimkets RA, Leach M
vv

DR WPI; 2000-602362/57.
D-DCDR: 33040075

[illegible]

PT useful for treati

XX
PS Claim 5; Page 1115; 5507pp; English.

AAC74446 to AAC77606 encode the prot

which represent the human ORFX open reading frames 1 to 31b1. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticoagulant; antiarthritic; immunosuppressant; immunostimulant; candidant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatologic; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 392 BP; 114 A; 101 C; 82 G; 95 T; 0 other;

Alignment Scores:

Pred. No.:	4,05e-25	Length:	392
Argument Scores:	Score: 656.00	Matches:	128
	Percent Similarity: 98.46%	Conservative:	0
	Best Local Similarity: 98.46%	Mismatches:	2
	Query Match: 9.52%	Indels:	0
	DB: 21	Gaps:	0

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:10:45 ; Search time 76 seconds
(without alignments)
5326.491 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence: 1 MODSIEASTSISQLRESY.....MESTMYSCSSRVVESDEL 1320

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+p2n.model -DEV=xlh
-Q=/cgn2_1/uspto.spool/US09818990/runat_26112002_093405_23250/app.query.fasta_1.1479
-DB-Issued Patents_NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.37 @runat_26112002_093405_23250 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5b_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6C_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	760	11.0	2244	US-09-354-151-1	Sequence 1, Appli
2	327	4.7	2793	US-08-795-868-13	Sequence 13, Appl
3	327	4.7	2793	US-09-303-069-13	Sequence 13, Appl
4	327	4.7	2793	US-09-134-250-13	Sequence 15, Appl
5	326	4.7	2614	US-08-795-868-15	Sequence 15, Appl
6	326	4.7	2614	US-09-303-069-15	Sequence 15, Appl
7	326	4.7	2614	US-09-134-250-15	Sequence 15, Appl
8	300	4.4	4608	US-09-041-886-24	Sequence 24, Appl
9	300	4.4	4608	PCR-US94-05277-1	Sequence 1, Appli
10	278.5	4.0	6814	US-09-484-970B-66	Sequence 66, Appl
11	278	4.0	4092	US-07-757-022B-51	Sequence 51, Appl
12	275	4.0	3420	US-07-757-022B-103	Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-354-151-1
; Sequence 1, Application US/09354151
; Patent No. 6458929
; GENERAL INFORMATION:
; APPLICANT: CARPEN, Olli
; APPLICANT: GRONHOLM, Mikaela
; APPLICANT: HEISKA, Leena
; APPLICANT: MYKANEN, Olli-Matti
; APPLICANT: SALMIKANGAS, Paula
; TITLE OF INVENTION: Myotilin, A No. 6458929a1 Actin-Organizing Protein
; FILE REFERENCE: 0933-0142P
; CURRENT APPLICATION NUMBER: US/09/354,151
; CURRENT FILING DATE: 1999-07-16
; EARLIER APPLICATION NUMBER: 60/093,169
; EARLIER FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2244
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (281)..(1774)
US-09-354-151-1

Alignment Scores:
Pred. No.: 1.98e-46 Length: 2244
Score: 760.00 Matches: 220
Percent Similarity: 45.95% Conservative: 98
Best Local Similarity: 31.79% Mismatches: 220
Query Match: 11.03% Indels: 155
DB: 4 Gaps: 19
US-09-818-990B-2 (1-1320) x US-09-354-151-1 (1-2244)

APPLICANT: Lee, Mu-En
APPLICANT: Hsieh, Chung-Ming
TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
TITLE OF INVENTION: AND STRIATED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,868
FILING DATE: 06-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,577
FILING DATE: 22-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/032001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 3...1983
OTHER INFORMATION:
US-08-795-868-13

Alignment Scores:
Pred. No.: 1.5e-14 Length: 2793
Score: 327.00 Matches: 165
Percent Similarity: 35.35% Conservative: 63
Best Local Similarity: 25.58% Mismatches: 254
Query Match: 4.75% Indels: 165
DB: 2 Gaps: 26

US-09-818-990B-2 (1-1320) x US-08-795-868-13 (1-2793)

Qy 35 AlaGluProSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 54
Db 409 AGCAGCCCAAGTCGAGCGCGGCCACCCGTCGGGCGACCCCGGCGCC-----455
Qy 55 GlyGlyGlnAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGlu 74
Db 456 -----TCGCAGGAAGAACTCGCGGCG 476
Qy 75 SerValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThr 94
Db 477 CCAGGCGAGCGTGGCGGCGCGCGCGCGCTGTTCCAG-----CAGAAAGCGCGCTCGCGT 530
Qy 95 GlnAlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGlu 114
Db 531 GACGAGCGACGCGT-----CAGCGCGCGCGCGCGCTCAGACCTCGAG 572
Qy 115 ProAsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGlu-----130
Db 573 CTGCGCTTCGCCAGGAGCTGGCGCGCTCCAGCTCGCGGAGGAGCTGGT 632

Qy 131 -----SerProGlnGluAlaLys 136
Db 633 CGTCGCGACAGTCCCTCGCGCGCCACCGCTCGAGCGTCCCGCATCCCTCGAGAGCCCGCG 692
Qy 137 ArgProGlnTyrCysSerGluThrGlnSerLysLysVal-----PheLeuAsnLysAla 154
Db 693 GAGCCCGCGCTCTTCTCTCGGCCCTCCACCCCAAGACATCGCGGGCGGTGAGCCCGCG 752
Qy 155 AlaAspPheIleGluGluLeuSerSerLysSerHisSerSerLysArgIleArg 174
Db 753 GCGCCCGCAGCGCCCTCTCCGAGCAGCGCGGAGAAGCG--GGGACGAGCCTGGGAGG 809
Qy 175 ProArgAlaCysLysAsnHisLysSerLysLeuGlu-----SerGlnAsnLys 190
Db 810 CCCAGGAGC---CGCGCGCGCGCGGAGCAGAGCGCGGGGAGGCCCGCCACAGAG 866
Qy 191 ValMetGlnGluAsnSerSerPheSerAspLeuSerGluArgArgGlu---ArgSer 209
Db 867 GTTAGCGCTCGGACCAATTCCCGCTGACCCGAGCAGAGCCATCCAGGAGTGCAGGAGC 926
Qy 210 SerValProIleProIle-----ProAlaAspThrArgAspAsn-----222
Db 927 CCTGTGCGCGCGCGCGCGCGCGATCCCGCAGAGCGCAGAGCAAGAACACCCCGCGTGG 986
Qy 222 -----222
Db 987 AAGCGGAGCGCGCGCGCGCGTTCCTGCGCTCGCGCCACCGCGCGCGCGCGCGTGGAG 1046
Qy 223 -----GluValAsnHisAlaLeuGlnGlnGlnGlnLys 234
Db 1047 GGCCTGCTGTACCCCGCAGACCTTGGAGAAGAAGAGCGGGCGCGCTGAGCAGAGAAGG 1106
Qy 235 ---ArgArgGluAlaGluGln-----AlaAla 242
Db 1107 CTTGCGAGAGCGCGGAGGAGGAGCGCTCCCTGGGGCGCGTGGACCGCGCGGCGCGCG 1166
Qy 243 SerGluAlaAlaGlyGlyAspThrThrProGlySer-----SerProSerSer 258
Db 1167 AGCAGGGCAAAAGTTCGCGCGCGCGCGCGCGCTCCCTGAGCTCGAGTCTTCGGATGAC 1226
Qy 259 LeuTyrTyr-----GluGluProLeuGlyGlnProArgPheThrGlnLysLeu 275
Db 1227 TCCTACGTGTCGCTGGAGAGAGCCCTA--GAGGCGCGCTGTGTGAGATCCCGCTG 1283
Qy 276 ArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValGlyIle 295
Db 1284 CAGAAATGTGTGTGGCACCAGGGCGCAGATGTGCTGCTCAAATGTATCATCACTGCCAAC 1343
Qy 296 ProProGlnValArgTyrTyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1344 CCGCGCGCGCAAGTGTCTTGGCACAAGATGGGTGAGCGCTGCGCAGCGAGGCGCGCTC 1403
Qy 316 HisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluAsp 335
Db 1404 CTCCTCGCGCTGAGGTGAGCGGCACACCTCTGCTCAGGAGGCGCGGCGAGCAGAT 1463
Qy 336 ThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGlu 355
Db 1464 GCGGGAGCTATATGGCCACCGCACCAACGAGCTGGGCCAGCACCTGTGCGCGCTCA 1523
Qy 356 IleTyrIleGluGlyValSerSerSerAspSerGluGlyAspProAsnLysGluGluMet 375
Db 1524 CTGACCTGAGACCGCGTGGGTCTACATCCCTTTTCAGCAGCCCGCATACCTCCGAGAG 1583
Qy 376 AsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAlaValIlePro 395
Db 1584 GAATACCTGAGCGCGCGCGAGAGGTTCACAGAGCTGGGAGAGC-----1628
Qy 396 ProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGln 415
Db 1629 -----TGCGCGCGCAACCCCGCCACCATGAAGCC 1655

QY 416 CysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaPro 435
Db 1656 AGTCCAGCCAGAACCGCGTCTTCTGACACTGGCTCAAAG-----GCACCCCCC 1706
QY 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGlu 455
Db 1707 ACCTTCAAGGTCTCACTTATGACACAGCTCAGTAAGAGAGGCCAAGATGTCATCATGAGC 1766
QY 456 CysArgValLysGlyAlaProSerProLysValGluTyrArgGluGlyThrLeuIle 475
Db 1767 ATCCCGGTGAGGGAGGCCACGAGCTGTGCTCTGCTGCTGAGAAACGCCAGCCCGGTG 1826
QY 476 GluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGlu 495
Db 1827 CGC-----CCAGAC-----CAGCGCGCTTTTGGGAGGAGGCT 1859
QY 496 Glu-----IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1860 GAGGTGGGCTGTGCGGCTCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1919
QY 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
Db 1920 ACTTGCAGCGGTCAAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1979
QY 534 GlyAsnGluAspLeuSerAsnAsnGlySerLeuHis----- 545
Db 1980 GGC-----GAGTG-AGCTCAGGGGGCCACCTGCGCTCCCGCGCTACCTCCGAGCCGC 2032
QY 546 -----SerAlaAsnSerThrThrAsnLeu-----AlaAlaIleGluProGln--- 559
Db 2033 GCCCTGTCTCAGGCACCTCTCGGACCTCGCTGTGTTTCACTGCTCTGCTGCTGCTGCTGCTGCT 2092
QY 560 -----ProSerProHisSerGluProProSer 569
Db 2093 CAGTGGCGGGCCGAGCCGCTCCAGCCTCCCTCCACCCCATGACAGCCGCCAGGGGG 2152
QY 570 ValGluGlnProPro 574
Db 2153 ATAGCCCATGGGCC 2167

RESULT 3

US-09-303-069-13
; Sequence 13, Application US/09303069A
; Patent No. 6350592
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; FILE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/039001
; CURRENT APPLICATION NUMBER: US/09/303,069A
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: US 09/134,250
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2793
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1985)
US-09-303-069-13

Alignment Scores:

Pred. No.:	1.5e-14	Length:	2793
Score:	327.00	Matches:	165
Percent Similarity:	35.35%	Conservative:	63
Best Local Similarity:	25.58%	Mismatches:	254
Query Match:	4.75%	Indels:	165
DB:	4	Gaps:	26

US-09-818-990b-2 (1-1320) x US-09-303-069-13 (1-2793)
QY 35 AlaGluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 54
Db 409 AGCAGCCCAAGTCGAGCGCGCGCACCGTGGGACCCCC- GGGGCC----- 455
QY 55 GlyGlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGlu 74
Db 456 -----TCGCAGGAAGAAGTGGGGCG 476
QY 75 SerValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThr 94
Db 477 CCAGCGACGCTGGCGGAGCGCGCGCTGTTCAG-----CAGAAAGCGCTCGCTG 530
QY 95 GlnAlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGlu 114
Db 531 GACGAGCCACGCGT-----CAGCGCAGCCCGCGCTCAGACCTCGAG 572
QY 115 ProAsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGlu----- 130
Db 573 CTGGCTTCGCGCAGGAGCTGGCGCGCATCCGCGCTCCACGTCGCGGGAGAGCTGGTG 632
QY 131 -----SerProGlnGlnAlaLys 136
Db 633 CGCTCGCAGAGTCCCTGCGCGCCACGCTGCAGCGTGCCTCCCTCGAGAGCCCGC 692
QY 137 ArgProGlnTyrCysSerGluThrGlnSerLysVal-----PheLeuAsnLysAla 154
Db 693 GAGCCCCGCTCTCTCGGCGCTCCACCCCAAGACATCGCGGCGCTGAGCCCGCC 752
QY 155 AlaAspPheIleGluLeuSerSerLeuPheLysSerHisSerSerLysArgIleArg 174
Db 753 GCCGCCAGCGCCCTCTCCGAGCAGCGCGAGAACGCG---GGGAGCAGACCTGGGAG 809
QY 175 ProArgAlaCysLysAsnHisLysSerLysLeuGlu-----SerGlnAsnLys 190
Db 810 CCCAGGAGC---CGCGGCGCGCGGAGGAGCAGACAGCGGGGAGGCGCCGACGAGGAG 866
QY 191 ValMetGlnGluAsnSerSerPheSerAspLeuSerGluArgArgGlu--ArgSer 209
Db 867 GTTAGGGCTCGGACCAATTCCTCGTGACCCGGAGCAGAGCCATCCAGGAGTGCAGGAGC 926
QY 210 SerValProIleProIle-----ProAlaAspThrArgAspAsn----- 222
Db 927 CCTGTGCGCGCCCGCGCGCGATCCCGCAGAGCAGGAGAAAGCACC-----CGGTGG 986
QY 222 ----- 222
Db 987 AAGCGGAGCCCGCGCGCAGCGCTGCGCTTCCTGCGCTGGGCGCACGCGCGGCTGGAG 1046
QY 223 -----GluValAsnHisAlaLeuGluGlnGlnAlaLys 234
Db 1047 GCGCTGCTGTACCCACAGACCTTGGAGAAACAGGCGGGGCGCTGAGGCAGAGAGAGG 1106
QY 235 ---ArgArgGluAlaGluGln-----AlaAla 242
Db 1107 CTTGCGAGAGGCGCGGAGGAGAGCGTCCCTGGGGCGCTCGGACCGCGAGGGCGCGC 1166
QY 243 SerGluAlaAlaGlyGlyAspThrProGlySer-----SerProSerSer 258
Db 1167 AGCCAGGCAAGGTCGCGGCGCGCGCCACCTCCCTGAGCTCGAGTCTTCGAGTAC 1226
QY 259 LeuTyrTyr-----GluGluProLeuGlyGlnProProArgPheThrGlnLysLeu 275
Db 1227 TCCTAGTGTCCGCTGGAGAGAGCCCTA---GAGGCGCTGTGTGTGAGATCCCGCTG 1283
QY 276 ArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValIcylIle 295
Db 1284 CAGAATGTGGTGGTGGCCACGAGGCGAGATGTGCTGCTCAAAATGTATCATCACTGCCAAC 1343
QY 296 ProProGlnValArgTyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1344 CCCCCGCCCCAAGTGTCTGCGCAAGGATGGTCAAGGCTGCGCAGCGAGCGGGCGGCTC 1403


```
QY 316 HistLeuValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAsp 335
    :::      |||      |||      |||      |||      |||      |||      |||
Db 1404 CTCCTCCGGCTGAGGTGAGCGGCACACCTGCTGCTCAGGGAGGCCAGGCGCAGCAT 1463
    :::      |||      |||      |||      |||      |||      |||      |||
QY 336 ThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGlu 355
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1464 GCCGGGAGCTATATGGCCACCGCACACAGGAGCTGGCGCCAGGCCACCTGTGCGCGCTCA 1523
    |||      |||      |||      |||      |||      |||      |||      |||
QY 356 IleTyrIleGluGlyValSerSerSerSerSerSerSerSerSerSerSerSerSerSer 375
    :::      |||      |||      |||      |||      |||      |||      |||
Db 1524 CTGACCTGAGACCGGTGGTCTTACATCCCTTTTACAGACGCCCATCATCTCCGAGCAG 1583
    |||      |||      |||      |||      |||      |||      |||      |||
QY 376 AsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAlaValIlePro 395
    :::      |||      |||      |||      |||      |||      |||      |||
Db 1584 GAATACCTGAGCCCCCAGAGGAGTCTCCAGAGCTGGGAGACC----- 1628
    |||      |||      |||      |||      |||      |||      |||      |||
QY 396 ProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGln 415
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1629 -----TGGCGCGGAACCCCAACCATGAGCC 1655
    |||      |||      |||      |||      |||      |||      |||      |||
QY 416 CysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaPro 435
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1656 AGTCCCAAGCCAGAACCCCGCTTCTTGACACTGGCTCCCAAG-----GCACCCCC 1706
    |||      |||      |||      |||      |||      |||      |||      |||
QY 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGlu 455
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1707 ACCTTCAAGTCTCACTATGGACCACTAGTAGAGAAGCCCAAGATGATCATCATGAGC 1766
    |||      |||      |||      |||      |||      |||      |||      |||
QY 456 CysArgValLysGlyAlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIle 475
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1767 ATCCGCTGAGGGGAGCCCAAGCCTGTGCTCTCTGCTGAGAAACCCGACGCCCGCTG 1826
    |||      |||      |||      |||      |||      |||      |||      |||
QY 476 GluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGlu 495
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1827 CGC-----CCAGAC-----CAGCGCGCTTTTGGGAGGAGGCT 1859
    |||      |||      |||      |||      |||      |||      |||      |||
QY 496 Glu-----IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1860 GAGGTGGGTGTGCGCGCTCGCATCTCGCTGCAGAGCGTGGCGATGCTGGTTCTAC 1919
    |||      |||      |||      |||      |||      |||      |||      |||
QY 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1920 ACTTGCAAGCGGTCAATAGTAGTGTGCTCGCAGTGCAGGCGCCGCTTGGAGGTCCGA 1979
    |||      |||      |||      |||      |||      |||      |||      |||
QY 534 GlyAsnGluAspLeuSerAsnAsnGlySerLeuHis----- 545
    |||      |||      |||      |||      |||      |||      |||      |||
Db 1980 GGC-----GAGTG-AGCTCAGGGGGCCACTGCGCTCCCGCCGCTACCTCCGAGCCGC 2032
    |||      |||      |||      |||      |||      |||      |||      |||
QY 546 -----SerAlaAsnSerThrThrAsnLeu-----AlaAlaIleGluProGln--- 559
    |||      |||      |||      |||      |||      |||      |||      |||
Db 2033 GCCCTGTCTCAGGCACCTCTCGACACCTGCTGTGTTTCACTGCTCTGCCACACACC 2092
    |||      |||      |||      |||      |||      |||      |||      |||
QY 560 -----ProSerProProHisSerGluProProSer 569
    |||      |||      |||      |||      |||      |||      |||      |||
Db 2093 CAGCTGCGGCGCCGAGCCGCTCCAGGCTCCCTCCCAACCCCATGAGCCCGCCAGGGG 2152
    |||      |||      |||      |||      |||      |||      |||      |||
QY 570 ValGluGlnProPro 574
    |||      |||      |||      |||      |||      |||      |||      |||
Db 2153 ATAGCCCATGGGCC 2167
    |||      |||      |||      |||      |||      |||      |||      |||

RESULT 4
US-09-134-250-13
; Sequence 13, Application US/09134250B
; Patent No. 6399753
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; FILE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/038001
; CURRENT APPLICATION NUMBER: US/09/134,250B
; CURRENT FILING DATE: 1998-08-14
```

```
; EARLIER APPLICATION NUMBER: US 08/795,868
; EARLIER FILING DATE: 1997-02-06
; EARLIER APPLICATION NUMBER: US 08/494,577
; EARLIER FILING DATE: 1995-06-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2793
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1985)
US-09-134-250-13

Alignment Scores:      1.5e-14      Length:      2793
Pred. No.:      327.00      Matches:      165
Score:      35.35%      Conservative:      63
Percent Similarity:      25.58%      Mismatches:      254
Best Local Similarity:      4.75%      Indels:      165
Query Match:      4      Gaps:      26
DB:

US-09-818-990b-2 (1-1320) x US-09-134-250-13 (1-2793)

QY 35 AlaGluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 54
    :::      |||      |||      |||      |||      |||      |||      |||
Db 409 AGCAGCCCAAGTCGGAGCGCGGCCACCGTGGGGCACCCCC-GGGGCC----- 455
    |||      |||      |||      |||      |||      |||      |||      |||
QY 55 GlyGlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGlu 74
    |||      |||      |||      |||      |||      |||      |||      |||
Db 456 -----TCGCAGGAAGAACTCGCGGCG 476
    |||      |||      |||      |||      |||      |||      |||      |||
QY 75 SerValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThr 94
    |||      |||      |||      |||      |||      |||      |||      |||
Db 477 CAGGCGAGCTGGCGCGCGCGCGCTGTTCAG-----CAGAAAGCGCGCTCGCTG 530
    |||      |||      |||      |||      |||      |||      |||      |||
QY 95 GlnAlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGlu 114
    |||      |||      |||      |||      |||      |||      |||      |||
Db 531 GACGAGCGCAGCGT-----CAGCGCACCCCGCGCTCAGACCTCGAG 572
    |||      |||      |||      |||      |||      |||      |||      |||
QY 115 ProAsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGlu----- 130
    |||      |||      |||      |||      |||      |||      |||      |||
Db 573 CTGCGCTTCGCCCGCAGGAGTGGCGCGCATCGCGCTCCACGTCGCGGAGGAGCTGGTG 632
    |||      |||      |||      |||      |||      |||      |||      |||
QY 131 -----SerProGlnGluAlaLys 136
    |||      |||      |||      |||      |||      |||      |||      |||
Db 633 CGCTCGCAGCAGTCCCTCGCGCGCCACGCTGCAGCGTGCCTCCCTCGAGAGCCCGCGC 692
    |||      |||      |||      |||      |||      |||      |||      |||
QY 137 ArgProGlnTyrCysSerGluThrGlnSerLysLysVal-----PheLeuAsnLysAla 154
    |||      |||      |||      |||      |||      |||      |||      |||
Db 693 GAGCCCGCGCTCTCTCTCGGCCCTCCACCCCAAGACATCGCGGGCGCTGAGCCCGCC 752
    |||      |||      |||      |||      |||      |||      |||      |||
QY 155 AlaAspPheIleGluLeuLeuSerSerLeuPheLysSerHisSerSerLysArgIleArg 174
    |||      |||      |||      |||      |||      |||      |||      |||
Db 753 GCCGCCACAGCGCCCTCTCCGAGCAGCGCGGAGAAGCG---GGGGACGACCTGGGAGG 809
    |||      |||      |||      |||      |||      |||      |||      |||
QY 175 ProArgAlaCysLysAsnHisLysSerLysLeuGlu-----SerGlnAsnLys 190
    |||      |||      |||      |||      |||      |||      |||      |||
Db 810 CCCAGGAGC---CGCGGGCGCGCGGAGCAGACAGCGGGGGAAGGCCCCAGCAGAGAG 866
    |||      |||      |||      |||      |||      |||      |||      |||
QY 191 ValMetGlnGluAsnSerSerSerPheSerAspLeuSerGluArgArgGlu---ArgSer 209
    |||      |||      |||      |||      |||      |||      |||      |||
Db 867 GTTAGGCGTGGGACCAATTCCCGCTGACCCGGAGCAGACCATCCAGAGTGCAGGAGC 926
    |||      |||      |||      |||      |||      |||      |||      |||
QY 210 SerValProIleProIle-----ProAlaAspThrArgAspAsn----- 222
    |||      |||      |||      |||      |||      |||      |||      |||
Db 927 CCTGTGCGCGCCCGCGCGCGATCCCCCAGAGGCCAGGACGAAAGACACCCCGGTGG 986
    |||      |||      |||      |||      |||      |||      |||      |||
QY 222 ----- 222
    |||      |||      |||      |||      |||      |||      |||      |||
Db 987 AAGCGGGAGCCCGCGCGCAGGCGGTGCGCTTCTTCCTGGGCGCACGCCGCGGCTGGAG 1046
    |||      |||      |||      |||      |||      |||      |||      |||
```

```
Qy 223 -----GluValAsnHisAlaLeuGluGlnGlnAlaLys 234
Db 1047 GCGCGTGTGTACCCAGACCTTGGAGAAAGACAGGGCGGCTGTGAGGCAGAGAGG 1106
Qy 235 ---ArgArgGluAlaGluGln-----AlaAla 242
Db 1107 CTTGCGCAGAGCGCGAGAGAGAGCGGTCCCTGGGGGCCCTGGGACCGCGGAGGGCGCGC 1166
Qy 243 SerGluAlaAlaGlyAspThrProGlySer-----SerProSerSer 258
Db 1167 AGCAGGGAAGAGTCCGGGGCGCGGCGCCACCTCCCTGTGAGCTCGAGTCTTCGGATGAC 1226
Qy 259 LeuTyrTyr-----GluGluProLeuGlyGlnProProArgPheThrGlnLysLeu 275
Db 1227 TCCTAGGTGTCCGCTGGAGAGAGAGCCCTA---GAGGCGCCCTGTGTTGAGATCCCGCTG 1283
Qy 276 ArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIle 295
Db 1284 CAGAATGTGTGTGGCACCAGGGCGAGATGTGCTCTCAATGTATCATCATCTGCCAAC 1343
Qy 296 ProProGlnValArgTyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1344 CCCCCGCCCAAGTGTCTGTCACAGATGGTACGGCTGCCAGCGGCGCGCTC 1403
Qy 316 HisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAsp 335
Db 1404 CTCCTCGGGGTGAGGTGAGCGGCACACCTGCTGTCTCAGGGAGCGGCAGGCAGAGAT 1463
Qy 336 ThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGlu 355
Db 1464 GCGGGAGCTATATGGCCACCGCCACCAACAGAGTGGGCGCCAGGCACCTGTGCGCGCTCA 1523
Qy 356 IleTyrIleGluGlyValSerSerAspSerGluGlyAspProAsnLysGluGluMet 375
Db 1524 CTGACCGTGACCGGTGGTGTCTACATCCCTTTCAGCAGCCCGCCACCTCCGAGCAG 1583
Qy 376 AsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAlaValIlePro 395
Db 1584 GAATACCTGAGCCCCCAGAGAGAGTTCACAGAGCTGGGGAGACC-----1628
Qy 396 ProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGln 415
Db 1629 -----TGGCGCGGAACCCACCATGAAGCCC 1655
Qy 416 CysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaPro 435
Db 1656 AGTCCAGCCAGACCGCGGTCTCTCAGACTGGCTCCAG-----GCACCCCCC 1706
Qy 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGlu 455
Db 1707 ACCTTCAAGGTCTCACTTATGGCAGCTCAGTAAGAGAAGGCCAAGATGTTCATCATGAGC 1766
Qy 456 CysArgValLysGlyAlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIle 475
Db 1767 ATCCCGGTGACGGGGAGCCCAAGCGTGTCTCTCGGTGAGAAACCGCCAGCCCGGTG 1826
Qy 476 GluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGlu 495
Db 1827 CGC-----CCAGAC-----CAGCGCGCTTTGCGGAGGAGGCT 1859
Qy 496 Glu-----IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1860 GAGGTGGGTGTGCGCGCTCGGATCCTGGCTGCAGAGCGTGGCGATGCTGTGTTTCTAC 1919
Qy 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
Db 1920 ACTTGCAAGCGGTCATAGATGGTGTCTCGGCAGTGCAGAGCGCGCTTGGAGGTCCGA 1979
Qy 534 GlyAsnGluAspLeuSerAsnAspGlySerLeuHis-----545
Db 1980 GGC-----GAGTG-AGCTAGGGGGGCACCTGTGCTCCCGCTCCCGCTCCGAGCGCG 2032
Qy 546 -----SerAlaAsnSerThrThrAsnLeu-----AlaAlaIleGluProGln--- 559
```

```
Db 2033 GCCCTGTCTCAGGACACCTCTCGGACCTGCTGTGTTTCACTGCTGCTGCCACAGACC 2092
Qy 560 -----ProSerProHisSerGluProProSer 569
Db 2093 CAGTGTCCGGCCCGAGACCCGTCGCCAGCTCCCTCCCAACCCCATGAGCCCCCAGGGGG 2152
Qy 570 ValGluGlnProPro 574
Db 2153 ATAGCCCATGGGCC 2167

RESULT 5
US-08-795-868-15
; Sequence 15, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; TITLE OF INVENTION: AND STRIATED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1803
; OTHER INFORMATION:
;
US-08-795-868-15

Alignment Scores:
Pred. No.: 1.6e-14 Length: 2614
Score: 326.00 Matches: 166
Percent Similarity: 35.10% Conservative: 79
Best Local Similarity: 23.78% Mismatches: 294
Query Match: 4.73% Indels: 160
DB: 2 Gaps: 25

US-09-818-990B-2 (1-1320) x US-08-795-868-15 (1-2614)
Qy 6 IleGluAlaSerThrSerIleSerGlnLeuArg---GluSerTyrLeuAlaGluThr 24
Db 421 ATCCGCGCATCTACGTGCGGGAGAGCTGCTGCGATTTCGACAGAGTCCCTGCTGCCACG 480
```

```
QY 25 ArgHisArgGlyAsnAsnGluArgSerArgAlaGluPro----- 37
D 481 CTGCAGCGCGCCCATCCCTCGGAGCGCGGAGCGCCCACTCTTCTCCCGGCTTCC 540
QY 38 -----SerSerAsnProCysHisPheGlySer 46
D 541 ACACCCCAAGACCTACGCGGTGTGAGCGCGGCTGCCACCCAGCGC-----CCGCGCT 591
QY 47 ProSerGlyAlaAlaGluGlyGlyGlnAspAspLeuProAspLeuSerAlaPhe 66
D 592 CTTAGTGTGCGGGCAAA---TCTGGGACGAGCCTGGGAGCGCCCGAAGCAGAGGCGCG 648
QY 67 LeuSerGlnGluLeuAspGluSerValAsnLeuAlaAraGluLeuAlaIleAsnTyrAsp 86
D 649 GTGGCAGGACTGAACCGGGGAAGCGCCGACGAGGAGATCAAGCTCGGACCAATTC 708
QY 87 ProLeuGluLysAlaAspGluThrGlnAlaAraGlyLysArgLeuSerProAspGlnMetLys 106
D 709 CCGCTAACCAAGAGAGAGAGCCATCCAGAGTGCAGG-----744
QY 107 HisSerProAsnLeuSerPheGluProAsnPheCysGlnAspAsnProArgSerProThr 126
D 745 ---AGCCCT-----GTGCGCCCTACACCGCGGATCCCGGAGAGCAGGACA 789
QY 127 SerSer-----LysGluSerProGlnGluAlaLysArgProGlnTyrCys 141
D 790 AAAGCCCTCCCGTCTCGCAAGCGGACACCCCTCTCAAGGGTGC-----837
QY 142 SerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGluLeu 161
D 838 -----TTTCTGCGCTGCCCGCCACTCCGGAGTGGAGAC---870
QY 162 SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis 181
D 871 -----TCTGTTCTGCCCGCAAACTTGGAGAGAAT 900
QY 182 LysSerLysLeuGluSerGlnAsnLysValMetGln-----GluAsnSerSerPhe 199
D 901 AGAGCGGACCGCGAGGTGAGAGAGCTTCGCAGAGGACCTGAGGAGGATGCCCTGTG 960
QY 200 SerAspLeuSerGluArgGluArgSerSerValProIleProAlaAspThr 219
D 961 GGGCCCTGGACCGCAGAGGACCGCAGC-----990
QY 220 ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGlu 239
D 991 -----CAAGGCAAGGTCCCGGTGCTCGG 1014
QY 240 GlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeu 259
D 1015 CCTACTTCCCGAGCTCGAGTCTCAGACGACTCTATGTGCTCGGTGG-----1065
QY 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu 279
D 1066 -----GAAGAGCCCTG---GAGGCACCGGTGTTTGAGATCCCTCTGCAGAAATATGGT 1116
QY 280 ValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGln 299
D 1117 GTGGCGCAGAGGTGACGTGCTACTTAAGTGTATCATCCCGCAACCCCGCCACCA 1176
QY 300 ValArgTyrTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln 319
D 1177 GTGTCTGGAAAGATGGTCCATCTTGCACAGCAGGCGTCTCTTCATCCCGGCT 1236
QY 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339
D 1237 GAAGGTGAACGGCACACACTGCTCTCAGAGAGGCGCCAGGCTGTGTGAGTGGAGGTAC 1296
QY 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
D 1297 ACAGCCACTGCCCAACCAAGTGGCCAAAGTACCTGTCTTCTTCACTGGCTGTGAGA 1356
```

```
QY 360 GlyValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGln 379
D 1357 CTTGGCGGCTCCACATCCCTTTTCAGCAGCCCATCATCTGTGATGAGGAGTACCTGAGC 1416
QY 380 LysProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValPro 399
D 1417 CCCCCAGAGAGTTCAGAGCCTGGGGAGAC-----1449
QY 400 GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerPro 419
D 1450 -----TGGCCCCGAAACCCCTACCATGAGCTCAGTCCAGCCAGCAG 1488
QY 420 ThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLys 439
D 1489 GATCATGATTCTCCGACTCTTCTTCCAAG-----GCACCCCAACGTTCAAGGTC 1539
QY 440 MetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLys 459
D 1540 TCACCTCATGACCAATCGGTGAGAGAAGTCAAGATCTCATATGAGCATCCCGTGTGAG 1599
QY 460 GlyAlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAspSerPro 479
D 1600 GGAGAGCCCAAGCTGTGGTTTCTGCTGAGGAATCGACAGCCCGTGC-----CCA 1653
QY 480 AspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGlu-----Ile 497
D 1654 GAC-----CAGCGGCGCTTTCAGAGGAGGCGGAGGTTGGGCTC 1692
QY 498 CysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAla 517
D 1693 TGCGCGTGTGAGGATCCTGGCTGTGAACGGGCGATCTGTGTTTCTACATGCAAGGCG 1752
QY 518 SerAsnLysTyrGlyThrValSerSerIleAlaGlnHisValArgGlyAsnGluAsp 537
D 1753 GTCAACAAATATGGCGCTCGCAGCTCGCAGCGCGCTGGAGGTCCGAGGCGAG-----1806
QY 538 LeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGlu 557
D 1807 ---TGAGCTCAGGGGGCCACCTGCGCTGCCCGCTACCTCCGAGCTGCACCCCTGTCT 1863
QY 558 ProGlnPro-----SerProHisSerGluPro-----567
D 1864 CAGCAGCTCTCGGACCTCGCTGTGTTTCTACTGCTCTCTGCCACAGCAGCCCGGCTC 1923
QY 568 -----ProSerValGluGlnProProLysProLysLeuGlyValLeuValAsnHis 585
D 1924 CCGCGCCGAGACTAGCCCATGTCCCTTCCCTCCCTCCTCCTCCTCCTCCTCCTCCT 1968
QY 586 AsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsnLeuProGluAsp 605
D 1969 ACAGCACCC-----TGGGGTAACCCACCGGCG 1995
QY 606 AspLysGlySerGluAlaSer-SerGluAlaGlyValValThrThrArgGlnThrArgPr 625
D 1996 CCCTGTGTGATCC-----TCCCTCCCAAGTGGATATGTGCTGTGCAGACAGGAGGCC 2049
QY 625 oAspSer**GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu---ProSerPh 644
D 2050 CCCAGAGG-ACTGAGTGTGGGAAGGATGGCCATGAGGGTCCCAAGCTCCCTCGGTC 2108
QY 644 eProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
D 2109 TCCCATAGGAGGATCCAGCGAGTGTGCTATGCTGCTGTACAGGCCACT 2160
```

RESULT 6

```
US-09-303-069-15
; Sequence 15, Application US/09303069A
; Patent No. 6350592
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
```


QY 568 -----ProSerValGluGlnProProLysProLysLeuGluGlyValLeuValAsnHis 585
Db 1924 GCGGGCGGACTAGCCATGCTCCCTTCCTCCCTAGCC-----CAT 1968
QY 586 AsnGluProSerSerArgIleGlyLeuArgValHisPheAsnLeuProGluAsp 605
Db 1969 ACAGCACCC-----TGGGGTAACCCACCGGGC 1995
QY 606 AspLysGlySerGluAlaSer-SerGluAlaGlyValValThrArgGlnThrArgPr 625
Db 1996 CCCTGTGGATCC-----TCCCTCCCAAGTGGATGCGGTGCGACACAGGAGGCC 2049
QY 625 oAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu---ProSerPh 644
Db 2050 CCCAGAAGG-ACTGAGTGTGGGAAGGATGGCCATGAGGGGTGCCAAGCTCCCTCGGTC 2108
QY 644 eProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
Db 2109 TCCCCATAGGAGCATCCAGCGAGTGCATGTGCTATGCTACAGGCCACT 2160

RESULT 7

US-09-134-250-15
; Sequence 15, Application US/09134250B
; Patent No. 6399753
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/038001
; CURRENT APPLICATION NUMBER: US/09/134, 250B
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: US 08/795,868
; EARLIER FILING DATE: 1997-02-06
; EARLIER APPLICATION NUMBER: US 08/494,577
; EARLIER FILING DATE: 1995-06-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1806)
US-09-134-250-15

Alignment Scores:
Pred. No.: Length: 1,6e-14 2614
Score: 326.00 Matches: 166
Percent Similarity: 35.10% Conservative: 79
Best Local Similarity: 23.78% Mismatches: 294
Query Match: 4.73% Indels: 160
DB: 4 Gaps: 25

US-09-818-990b-2 (1-1320) x US-09-134-250-15 (1-2614)

QY 6 IleGluAlaSerThrSerLysLeuLeuArg---GluSerTyrLeuAlaGluThr 24
Db 421 ATCCGGCGGATCTACGTGCGGGAGGAGCTGGTGGTTCGACACGAGTCCCTGCGTGCACG 480
QY 25 ArgHisArgGlyAsnAsnGluArgSerArgAlaGluPro----- 37
Db 481 CTCGAGCGGCGCCCATCCCTCGGGAGCGCGGAGCGCCCGCCCATCTCTCTCCCGGCTTCC 540
QY 38 -----SerSerAsnProCysHisPheGlySer 46
Db 541 ACACCAAGACCTCACGGGGTGTGAGCGCGCGGTGCCACCGCCG-----CCGCGCT 591
QY 47 ProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeuProAspLeuSerAlaPhe 66
Db 592 CCTAGTGGTGGCGCAAAA---TCTGGGAGCGAGCGCTGGGAGGCGCCCGAAGCAGAGGGCGG 648

QY 67 LeuSerGlnGluLeuLeuAspGluSerValAsnLeuAlaArgLeuAlaLeuAsnTyrAsp 86
Db 649 GTGGGAGGACTGAACCGGGGAGAGCCCGCACGAGAGATCAAGCGTCGGGACCAATTC 708
QY 87 ProLeuGluLysAlaAspGluThrGlnAlaArgLysLeuSerProAspGlnMetLys 106
Db 709 CCCTTAACCAAGGAGAGACCATCCAGGAGTGCAGG----- 744
QY 107 HisSerProAsnLeuSerPheGluProAsnPheCysGlnAspAsnProArgSerProThr 126
Db 745 ---AGCCCT-----GTCCCGCCCTACACCGCGGATCCCGCGAGACGAGACA 789
QY 127 SerSer-----LysGluSerProGlnAlaLysArgProGlnTyrCys 141
Db 790 AAGCCCTCCCGTCGTCGACGAGCGGAACCCCTGCTCAACGCGTGCCT----- 837
QY 142 SerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGluLeu 161
Db 838 -----TTTCTGCCCTGGGCCACTCCGGGAGTGGAGGAC--- 870
QY 162 SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 871 -----TCTGTTTGCCTCCCAACCTTGGGAGAAGAT 900
QY 182 LysSerLysLeuGluSerGlnAsnLysValMetGln-----GluAsnSerSerPhe 199
Db 901 AGACGGGAGCCCGAGCTGAGAGAGCTTCGACGAGGACCTGAGGAGGATGCCCTGG 960
QY 200 SerAspLeuSerGluArgGluArgSerSerValProIleProIleProAlaAspThr 219
Db 961 GGGCCCTGGGACCGCAGAGGAGCCCGCAGC----- 990
QY 220 ArgAspAsnGluValAlaAsnHisAlaLeuGluGlnGlnGluAlaLysArgGluAlaGlu 239
Db 991 -----CAAGCAAAAGTCCCGCTGCTCGG 1014
QY 240 GlnAlaAlaSerGluAlaAlaGlyGlyAspThrProGlySerSerProSerSerLeu 259
Db 1015 CCTACTTCCCGGAGCTCGAGTCCCTCAGAGACTCTATGTGTCGCTGGG----- 1065
QY 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerGlu 279
Db 1066 -----GAAGAGCCCTG---GAGGACCCGCTGTTGAGATCCCTCTGCAAGATATGTTG 1116
QY 280 ValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGln 299
Db 1117 GTGGCGGAGGAGCTGACGTGCTACTTAAGTGTATCATCACCGCCCAACCCCGCCCA 1176
QY 300 ValArgTyrPyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln 319
Db 1177 GTGTCCTGGAAAAGGATGGTCCATGTCACAGCGAGGCTGCTCTCTCATCCGGGCT 1236
QY 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339
Db 1237 GAAGGTGAACGGCACACACTGCTGCTCAGAGAGGCCCGAGCTGCTGATGTTGGAGTAC 1296
QY 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
Db 1297 ACAGCCACTGCCACCAACGAACTGGGCAAGCTACCTGCTCTTCTCAGCTGGTGAGA 1356
QY 360 GlyValSerSerSerSerGluGlyAspProAsnLysGluGluMetAsnArgIleGln 379
Db 1357 CCTGGCGGCTCCATCCCTTTTCAGCAGCGCCCATCATCTCTGATGAGGAGTACCTGAGC 1416
QY 380 LysProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValPro 399
Db 1417 CCCCCAGAGAGTTCACAGACCTGGGGAGACC----- 1449
QY 400 GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerPro 419
Db 1450 -----TGGCCCCGAAACCCCTACCATGAAGCTCAGTCCCGAGCCAG 1488
QY 420 ThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLys 439

```
Db 1489 GATCATGATTCCTCCGACTCTTCTTCCAAG-----GCACCCCAACGTTTCAAGGTC 1539
Qy 440 MetLeuGlnAsnLeuSerAlaSerGluGlnLeuValPheGluCysArgValLys 459
Db 1540 TCACATGAGACCAATCGGTGAGAGAGGTCAAGATGTCATATGAGCATCCGTTGTGAG 1599
Qy 460 GlyAlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerPro 479
Db 1600 GGAGAGCCCAAGCCTGTGTTCTTGCTGAGGAATCGACAGCCCGTGCGC-----CCA 1653
Qy 480 AspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGlu-----Ile 497
Db 1654 GAC-----CAGCGCGCTTTCAGAGGAGCGCGAGGGTGGGCTC 1692
Qy 498 CysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAla 517
Db 1693 TGCCTGTGAGATGCTGCGTGTGAACGGGGCGATGCTGTTCTACACATGCAAGGCG 1752
Qy 518 SerAsnLysTyrGlyThrValSerIleAlaGlnLeuHisValArgGlyAsnGluAsp 537
Db 1753 GTCAACGAATATGCTGCGCTGCGCACTGCGAGCGCGCTCGGAGTCCAGGGCGAG----- 1806
Qy 538 LeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGlu 557
Db 1807 ---TGAGCTCAGGGGGCCACCTGCGCTGCGCGCTACCCCTCCAGAGTGCACCCCTGTCT 1863
Qy 558 ProGlnPro-----SerProHisSerGluPro----- 567
Db 1864 CAGGCACTCTCGGACCTCGCTGTGTTTTCACGTGCTCTGCTGCCACAGACCCGCGGCTC 1923
Qy 568 -----ProSerValGluGlnProProLysProLysLeuGluGlyValLeuValAsnHis 585
Db 1924 GCCGCGCGGACTAGCCCATGCTCCCTTCCCTCCCTAGCC-----CAT 1968
Qy 586 AsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsnLeuProGluAsp 605
Db 1969 ACAGCACCC-----TGGGCTAACCCACCGCGGC 1995
Qy 606 AspLysGlySerGluAlaSer-SerGluAlaGlyValValThrThrArgGlnThrArgPr 625
Db 1996 CCCTGTGATCC-----TCCCTCCCAAGTGGATGTGTGCTGTGCAGACAGGAGGCC 2049
Qy 625 oAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu---ProSerPh 644
Db 2050 CCCAGAAGG-ACTGAGTGTGGGAAGGATGGCCATGAGGGGTCCCAAGTCCCTCGGTC 2108
Qy 644 eProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
Db 2109 TCCCATAGGAGGATCCAGCGAGTGTGCTGTATGCTGTATGCTACAGGCCACT 2160
```

RESULT 8

```
US-09-041-886-24
; Sequence 24, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic peptides, dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4342
; US-09-041-886-24

Alignment Scores:
Pred. No.: 3,03e-12 Length: 4608
Score: 300.00 Matches: 108
Percent Similarity: 41.13% Conservative: 45
Best Local Similarity: 29.03% Mismatches: 135
Query Match: 4.35% Indels: 84
DB: 15 Gaps: 15

US-09-818-990b-2 (1-1320) x US-09-041-886-24 (1-4608)
Qy 959 GlySerProValThrPheThrCysLysIleValGlyLeProValProLysValTyrTrp 978
Db 460 GGAGACACAGCTGTACTCAAGTGTGAAGTCTGGGAGGCCATGCCAACATCCACTGG 519
Qy 979 PheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluGlu 998
Db 520 CAGNAGAACCAACAGACCTGACTCCA-----ATCCCCAGGT 555
Qy 999 AspGly-----ThrCysSerLeuHisIleGluSerThrThrSerAsp 1012
Db 556 GACTCCCGAGTGTGTGCTTGCCTCTGGAGCATTCAGATCCAGCCGACTCCAACCGGG 615
Qy 1013 AspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGly 1032
Db 616 GACATTGGAATTTACCGATGCTCAGCTCGAAATCCAGCCAGC-----TCAAGAACAGGA 669
Qy 1033 HisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis 1052
Db 670 AAT-----GAAGCAGAAGTCAGATTTTATCAGATCCAGGACTGCAT 711
Qy 1053 ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArg 1072
Db 712 AGA-----CAG 717
Qy 1073 ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu 1092
Db 718 CTGTATTTCTGCAAGACCATCAATGTAGTACCCATTGAAGGAAAGATGCTGCTCTG 777
Qy 1093 AspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlnPro 1112
Db 778 GAATGTGTGTCTGGCTATCTCCACCAAGTTTACCTGG---TTACGAGCGGAGGAA 834
Qy 1113 ValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1132
Db 835 GTCATCCAACCTCAGGCTCAAAAAGTATTTCTTTA-----TTGGGTGGAAGCAACTGCTT 888
Qy 1133 IleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThr 1152
Db 889 ATCTCCAATCTGACAGATGATGACAGTGAATGTATCATCTGTGTGTACATATAAAAT 948
```

```
QY 1153 GlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaPro 1172
Db 949 GAGAAATATTAGTCCCTGCGAGAGTCACAGTCTTGTT-----CCGCCA 993
QY 1173 ValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGlu 1192
Db 994 TGGTTTTTAATCATCTTCCCAACCTGTATGCTATGAAAGCAGCATGATATTGAGTTTGAA 1053
QY 1193 CysArgValIleGlyMetProProValPheTyrTriPlyLysLysAspAsnGlu---Thr 1211
Db 1054 TGTACAGTCTCTGGAAGCCTGTGCCACTGTGAATGGATGAGATGAGATGTCG 1113
QY 1212 IleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeu 1231
Db 1114 ATTCCT-----AGTGATTATTTTCAGATAGTGGGAGGAGCAACTTA 1155
QY 1232 LeuIleGlnProAlaLysLysSerAspAlaGlyTrpThrLeuSerAlaLysAsnGlu 1251
Db 1156 CGATACTTGGGTGGTGAGTCCAGTGTCTCCCTTCGGCTCCAGAGATGGTCCCT 1311
QY 1252 AlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIle 1271
Db 1216 GCTGGAATCCCGACAGCAGTGCACAGCTC-----ATTGTC 1251
QY 1272 Pro-----ProMetSerValArgProSer----- 1280
Db 1252 CCTAAGCCTGCAATCCCAAGCTCCAGTGTCTCCCTTCGGCTCCAGAGATGGTCCCT 1311
QY 1281 -----GlySerArgTyrGlySerLeuThr-----SerLysGly 1291
Db 1312 GTCTTGTTTCCAGCGCATTTGTCCGTCTCAGCTGGCGCCACCTGCAGAGGAAAGGG 1371
QY 1292 ---LeuAspIlePheSerAlaPheSerSerMetGlu 1302
Db 1372 AACATTCAAACTTCAGGCTCTTTTCTCCAGAGAA 1407

RESULT 9
PCT-US94-05277-1
; Sequence 1, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zabrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 1:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 18q21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4342
PCT-US94-05277-1

Alignment Scores: 3.03e-12 Length: 4608
Pred. No.: 300.00 Matches: 108
Score: 41.13% Conservative: 45
Percent Similarity: 29.03% Mismatches: 135
Best Local Similarity: 4.35% Indels: 84
Query Match: 5 Gaps: 15
DB:

US-09-818-990B-2 (1-1320) x PCT-US94-05277-1 (1-4608)
QY 959 GlySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrp 978
Db 460 GGAGACACAGTGTCTACTCAAGTGTGAAGTCTGGGAGCCCATGCCAACATCCCACTGG 519
QY 979 PheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGly 998
Db 520 CAGAGAACCACACAGACTGACTCCA-----ATCCCAGGT 555
QY 999 AspGly-----ThrCysSerLeuHisIleGluSerThrThrSerAsp 1012
Db 556 GACTCCCGAGTGGTGTCTTGCCTCTGGAGCATTCGAGATCAGCCGACTCCAACCGGG 615
QY 1013 AspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGly 1032
Db 616 GACATTGGGAATTTACCGATGCTCAGCTCGAAATCCAGCCAGC-----TCAAGAACAGGA 669
QY 1033 HisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis 1052
Db 670 AAT-----GAAGCAGAACTCAGAAATTTATCAGATCCAGGACTGCAT 711
QY 1053 ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArg 1072
Db 712 AGA-----CAG 717
QY 1073 ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu 1092
Db 718 CTGTATTTCTCAAGACACCATCAATGTAGTACCCATTTGAAGGAAAGATGCTGCTCTG 777
QY 1093 AspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnPro 1112
Db 778 GAATGTGTCTTCTGGCTATCTCCACCAAGTTTACCTGG---TTACAGCGCGAGGAA 834
QY 1113 ValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeu 1132
Db 835 GTCATCCAACTCAGGTCTAAAAAGTATTCCTTA-----TTGGGTGGAAGCAACTGCTCT 888
QY 1133 IleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThr 1152
Db 889 ATCTCCAATGTGACAGATGATGACAGTGGAAATGATATACCTGTGTGTCACATATAAAAT 948
QY 1153 GlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaPro 1172
Db 949 GAGAAATATTAGTCCCTGCGAGAGTCACAGTCTTGTT-----CCGCCA 993
QY 1173 ValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGlu 1192
Db 1192
```


Db	3303	GCTG-----GACGAGAACTTCACGGCCCATTTGCACAAAGATCACTCCCCCT	3343
Qy	939	rgILyScysIleAlaProIlePheAspLysArgLeuLysHisPheArg---ValThrGI	958
Db	3348	TCACAAAGCTTTC-----TTCTCTCCCTCCGGATTGTGAATGA	3386
Qy	958	uGly-SerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrT	978
Db	3387	GGCGGCATCGATCCGCTTCTCAGGGGGCTGTTCGGG-----T	3425
Qy	978	rPheLysaspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluG	998
Db	3426	GG-----CGGGAAATCGTGTCCCTCGCAGC-----TGTGAAACACGGAGCTCAC	3473
Qy	998	lyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspClyAsnTyrT	1018
Db	3474	GGAGCGGCTGTTCTCATGGCACACAGGTGGCTC-----TGGG	3512
Qy	1018	hrIleMetAlaAlaAsnProGInGlyArgIleSerCysSerGlyHisLeuMetValGlnS	1038
Db	3513	CCTGGCGGCATCAACATCCAGCGGGCGGACACGGATCCAC-----	3559
Qy	1038	erLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgV	1058
Db	3560	-----CCTACACAGACTACAGGGTGCTACTGCAATCTATCGCGCGCACACAGTTCAGGA	3614
Qy	1058	alGInGluArgaspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnA	1078
Db	3615	CCTGAAAATGAGATTAAAAACCTTGAGATCCGGG-----AGAAACT	3656
Qy	1078	laProGlyAspMetValAlaHisGluGlyArgLeuCys-----ArgL	1092
Db	3657	GAAAAGGTGATGGCTCGACACTCAACATCGACCTGTTCCGGCGCTCGTGGTGAGGA	3716
Qy	1092	euAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyClnP	1112
Db	3717	CCTGGTGCTTCGCGAGCGGCTGGGCGCCACCTGATGTGCTTC-----	3760
Qy	1112	roValLeuProAspAlaSerHisLysMetLeuValArgGlu-----ThrGlyV	1128
Db	3761	-----TCAGCACACAGTTCAAGCGCTCGCAGATGGGACAGAGTTGTGGTA	3806
Qy	1128	alHisSerLeuLeuIleAspPro-----LeuThrGlnArgAspA	1141
Db	3807	TGAGAACCTCGGGGTCTTCTCCCGCCGCGAGCTGACTCAGATCAAGCAGACGCTGCTGGC	3866
Qy	1141	laGlyThrTyrLysCysIleAlaThrAsnLysThrGly-----GlnAsnSerP	1157
Db	3867	CAGAGTCTATCGCACACGCGGACAGCATCACCCGGGTGCGAGCAGCAGTGTTCAGGT	3926
Qy	1157	heSerLeuGluLeuSerValAlaLysGluValLysLysAlaProValIleLeuGluL	1177
Db	3927	GGCGGAGTCCCTCAGCGGTACGGCA---GCTGTACGAGATCCCA-----GGGT	3974
Qy	1177	ysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleG	1197
Db	3975	GGACCTCCGGGTGTGGC-----AGGACTGCTGTGAAGACTGTAG	4013
Qy	1197	ly-----MetProProProValPheTyrTrpLysLysAspAsnGluThrI	1212
Db	4014	GACACGGGGCAGTTCAATAGCTCTTTTCCTATC-----	4045
Qy	1212	leProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuL	1232
Db	4046	-----ATTTCGAGGACGAGCGGTCTCTTGAGTTTCAGTACCAGGA	4085
Qy	1232	euIleGlnProAlaLysLysSerAspAlaGlyTyrTrpThrLeuSerAlaLysAsnGluA	1252
Db	4086	GGACAAGCCGACCAAGAAAACAAAGACCAGGAAATACCCAGTG-----TTGGGAGACA	4139
Qy	1252	laGlylleValSerCysThrAla---ArgLeuAspIleTyrAlaGlnTrpHisHisGlnI	1271
Db	4140	GGGGGAACATCTCAGCAACAGACACCTCAGCTTCAGACACGGCTCAGATGCAT-----	4192

```

Qy 1271 leProPrometSerValArgProSerGlySerLeuThrSerLysG 1291
Db 4193 -----CTGGGACAAATGACTTTCAGAGA 4214
Qy 1291 lyLeu 1292
Db 4215 GTTTC 4219
RESULT 11
US-07-757-022B-51
: Sequence 51, Application US/07757022B
: Patent No. 6433142
: GENERAL INFORMATION:
: APPLICANT: Gesner, Thomas G.
: APPLICANT: Clark, Stephen C.
: APPLICANT: Turner, Katherine
: APPLICANT: Hewick, Rodney M.
: TITLE OF INVENTION: Megakaryocyte Stimulating Factors
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07757.022B

```

Alignment Scores:		
Pred. No.:	1.04e-10	4092
Score:	278.00	239
Percent Similarity:	31.41%	115
Best Local Similarity:	21.21%	397
Query Match:	4.03%	379
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	

[illegible]

```
Db 1469 CCACCACCTACCAAGGAGCGCTGCACCCACCACTACCAAGTCTGCACCCACCACTCCCAAGG 1528
Qy 753 hrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerValGlnThrL 773
Db 1529 AGCCTTCACCCACCAACCAAG-----1551
Qy 773 ysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProProGlyProThrGluProT 793
Db 1552 -----GAGCCTGCACCCACCACTCCCAAGGAGCGCTG 1582
Qy 793 hr-ProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysVal 812
Db 1583 CACCACCAACCC---CCAGAAGCGCTGCGC-----CAACTACCCCAAGGAGCGCTG 1630
Qy 813 SerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSerSer 832
Db 1631 CACCCA---CAACTC---CCAGGAACTGCACCCACCAACCAAGAGCCTGCACCCA 1684
Qy 833 ValLeuPro---SerLeuProAlaIleProProThrAsnAla---Met***LeuProArg 850
Db 1685 CGCTCCCAAGAGCGCTGCCCAACTACCCCAAGGAGACTGCACCCACCAACCCCAAGA 1744
Qy 851 SerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnPro 870
Db 1745 AGCTCAGCCGCCACCAACCCCGAGAAGCTGCGCACCCACCAACCCCTGAGAAGCGCGCACCA 1804
Qy 871 -----ValAsnAspAsnIleArgGluThrLysAsnAlaValIleArg 885
Db 1805 CCACCCCTGAGAGCTGCGCACCCACCCCTGAGAGCGCCACACCCACCACTGAGG 1864
Qy 886 AspLeuGlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGluTyrLysIle 905
Db 1865 AGCCTGCTCCCACTCCCA-----1885
Qy 906 SerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProVal 925
Db 1886 -----AGGACCGCGCTCCCAACACCCCTAAGGAGCGCTCTCCAACCT-ACCCCTAAG 1935
Qy 926 AspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysIleAlaPro 945
Db 1936 GAGCTGCTCCCACTACCCCTAAGGAGCGCTCTCCAACCTACCCCTAAGGAGACTGCTCCA 1995
Qy 946 IlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
Db 1996 ACTACCCCTAAAGGAGCTGCTCCAACCTACCCCTCAAGGAACCTGCACCCCACTACT-----2049
Qy 966 CysLysIleValGlyIleProValProLysValTyrTyrPheLysAspGlyLysGlnIle 985
Db 2050 -----CCCAAGAAGCGCTGCCCCCAAG-----2070
Qy 986 SerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHis 1005
Db 2071 -----GAGCTTGCAACCCACCAACCAAGGAG 2097
Qy 1006 IleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGln 1025
Db 2098 CCCATCCACCACTCTGACAAGCCGCTCCAACCTACCCCTCAAGGAACCTGCACCC 2142
Qy 1026 GlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArgLe 1045
Db 2143 GGGAGCTGCTCAACTACCCCTAAGGAGCGCTGCTCCAA---CTACCCCTAAGGAGCGCTGCT 2199
Qy 1045 uThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgAspLysGluPr 1065
Db 2200 CCAA-----CTACCCCTAAGGAGCTGCTCCAACCTACCCCTCAAGGAACCTGCACCC 2250
Qy 1065 oLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHi 1085
Db 2251 ACTACTCCCAAGAGCGCTGCCCAAGGAGCTTGCAACCAACCAAGGAGGCGCCACA 2310
Qy 1085 sGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProProGluLeuTh 1105
Db 2311 T-----CCACCACCTCTGAC 2325
```

```
Qy 1105 rTrpLeuLeu-----AsnGlyGlnPr 1112
Db 2326 AAGCCTGCTCCAACCTACACCTAAGGAGACTGCTCCAACCTACCCCAAGGAGCCTGCACCC 2385
Qy 1112 oValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLe 1132
Db 2386 ACTACCCCAAGAGAGCGCTGCTCCAACCTACTCTCTGAGACAC-----CT 2427
Qy 1132 uIleAspProLeu 1136
Db 2428 CCTCCAACCACTT 2440
RESULT 12
US-07-757-022B-103
: Sequence 103, Application US/07757022B
: Patent No. 6433142
: GENERAL INFORMATION:
: APPLICANT: Gesner, Thomas G.
: APPLICANT: Clark, Stephen C.
: APPLICANT: Turner, Katherine
: APPLICANT: Hewick, Rodney M.
: TITLE OF INVENTION: Megakaryocyte Stimulating Factors
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07757,022B
: FILING DATE: 19910910
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/643,502
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/546,114
: FILING DATE: 29-JUN-1990.
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/457,196
: FILING DATE: 29-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/390,901
: FILING DATE: 08-AUG-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Cserr, Luann
: REGISTRATION NUMBER: 31,822
: REFERENCE/DOCKET NUMBER: GI 5190
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)876-1170
: TELEFAX: (617)876-5851
: INFORMATION FOR SEQ ID NO: 103:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3420 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3420
US-07-757-022B-103
Alignment Scores: 1.32e-10 Length: 3420
Pred. No.:
```

Score: 275.00 Matches: 229
Percent Similarity: 31.03% Conservative: 103
Best Local Similarity: 21.40% Mismatches: 376
Query Match: 3.99% Indels: 365
DB: 4 Gaps: 43

US-09-818-990B-2 (1-1320) x US-07-757-022B-103 (1-3420)

```
QY 116 AsnPheCysGlnAsp-----AsnProArgSerProThrSerLysGluSerProGln 133
Db 304 AGTTCTGTGCAGAGTGCATATCCACATCACCACCATCTTCAAAGAAGACCTCCA 363
QY 134 GluAlaLysArgProGlnTyrCysSerGluThrGlnSerLysLysValPheLeuAsnLys 153
Db 364 -----CCTTCAGGAGCATCTCAAACC----- 384
QY 154 AlaAlaAspPheIleGluGluLeuSerSerLeuPheLysSerHisSerLysArgIle 173
Db 385 -----ATCAATCAACACCAACCAACGTTTCAACC 411
QY 174 ArgProArgAlaCysLysAsnHisLysSerLysLeuGluSerGlnAsn----- 189
Db 412 AAACCCCAACAAAGAGAGACAGTAAGAGTTATAGAAACAGAGAAATAACAGAGAA 471
QY 190 -----LysValMetGlnGluAsnSerSerSerPheSerAsp----- 201
Db 472 CATTTCTGTTTGAATAAATCAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 531
QY 202 -----LeuSerGluArgGluArgGluArgSerSerValProIleProIlePro 216
Db 532 ACATTTTGGAAAAATCAAGTCTTCCAAAATTCAGCTGCTAATAGAGAAATTACAGAAGAA 591
QY 217 AlaAspThrArgAspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArg 236
Db 592 CTCAAAGTAAAGATACAGAGAGACAGACTAAAGAAACCTACCCCAACACCA 651
QY 237 GluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAsp-----ThrThrProGly 253
Db 652 GTTCTAGATGAAGCTGGAAGTGGACATGGTACATTCAGGTCAAGTCAACACTCTCTGAC 711
QY 254 SerSerProSerSerLeuTyrTyrGluGluProLeuGlyGlnProProArgPheThrGln 273
Db 712 ACGTCTACCAACCCAA-----CACAATAAAGTACAGCATCTCCCAAGATCACACA 762
QY 274 LysLeuArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleVal 293
Db 763 GCAAAACCAATAATCCAGACC----- 786
QY 294 GlyIleProProGlnValArgTyrTyrCysGluGlyLysGluLeuGluAsnSerPro 313
Db 787 AGTCTTCCACCT----- 798
QY 314 AspileHisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGlu 333
Db 799 -----AAT 801
QY 334 GluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSer 353
Db 802 TCTGATACATCTAAGAGACCTCTTGCAGTGAATAAAGACACACAGATTGAACCTAAA 861
QY 354 AlaGluIleTyrIleGluGlyValIserSerSerAsp---SerGluGlyAspProAsnLys 372
Db 862 GAAACTACTACAACAAATAACACACTTCAACTGATGAAAAGAGAGACTACTTCCGCT 921
QY 373 GluGluMetAsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAla 392
Db 922 AAAGACACAAAGATATAGAGAAACATCTGCTAAAGATTTAGCACCCACATCTAAAGTG 981
QY 393 ValIleProProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThr 412
Db 982 CTGCTAAACCT-----ACACCCCAAGCTGAACACTACAACCAAGCCCTGCTCTCACCAC 1038
QY 413 IleGlnGlnCysGlnSerProThrAsnTyrLeuGlnGlnGlyLeuAspGlyLysProIleIle 432
```

```
Db 1039 CCCAAGGAG---CCACGCGCCACCATCTCCCAAGGAGCGCTGCATCTACCAACCCAAAGAG 1095
QY 433 AlaAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuVal 452
Db 1096 CCCACACCTACC----- 1107
QY 453 ValPheGluCysArgValLysGlyAlaProSer-ProLysValGluTrpTyrArgGluGlu 472
Db 1108 -----ACCATCAAGTCTGCACCCACCAACC----- 1132
QY 472 yThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetal 492
Db 1133 -----CCAAGGAGCGCTGCA 1146
QY 492 aGluPro-GluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyC 512
Db 1147 CCCACCAACCAAGTCTGCAC----- 1169
QY 512 ysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisV 532
Db 1170 -----CACCACT 1176
QY 532 alArgGlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrA 552
Db 1177 CCCAAGGAGCGCTGCACCCACCAACCAAGGAGC-CTGCAC----- 1216
QY 552 snLeuAlaAlaIleGluProGlnProSerProProHisSerGluProProSerValGluG 572
Db 1217 -----CCACCACTCCCAAGGAGCGTGCACCCACCAACCAAGGAGCGCTGC 1262
QY 572 lnProLysProLysLeuGluGlyValLeuValAsnHisAsnGluProArgSerSers 592
Db 1263 ACCCACCAACCAAGTCTGCACCCACCTC-----CCAAGGAGC----- 1303
QY 592 erArgIleGlyLeuArgValHisPheAsnLeuProGluAspAspLysGlySerGluAla 612
Db 1304 -----CTGCACCC-ACCACCCCAAG---AAGCTGCCCCCAACTACCC 1342
QY 612 erSerGluAlaGlyValValThrArgGlnThrArgProAspSer***GlnGluArgP 632
Db 1343 CCAAGGAGCGCTGCACCCACCACTCCCAAGGAGCGCTACACCCACCACTCCCAAGGAG---- 1398
QY 632 heAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProProV 652
Db 1399 --CCTGCACCCACCAAGGAGCGTGCACCCACCACTCCC---AAAGAGCTCTGCACCCA 1453
QY 652 alLeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeu 672
Db 1454 CTGCCCCCAAGAGCGCTGCCCAACTACCCCC----- 1485
QY 672 euGluGlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe*** 692
Db 1486 -----AAGGAGCGTGCACCCACCACTCCCAAGGAG----- 1515
QY 692 etThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysA 712
Db 1516 -----CTGCACCCACCACTCCCAAGGAG-----AAGG 1537
QY 712 laProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnT 732
Db 1538 AGCCTTCCACCCACCACT-----CCAAGGAGCGCTGCACCCACCACTCCCAAGGAG 1582
QY 732 hrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer-----SerT 750
Db 1583 AGTCTGCACCCACCACTACCAAGGAGCGTGCACCCACCACTACCAAGTCTGCACCCACCA 1642
QY 750 hrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerV 770
Db 1643 CTCCCAAGGAGCGTTCACCCACCACTCCCAAG----- 1674
QY 770 alGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyProT 790
```

Db	1675	-----GAGCTGCACCCACCACTCCCA	1696
Qy	790	hrGluProThr-ProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnPro	809
Db	1697	AGGAGCTGCACCCACCACTCCCAAGAACGCTGCCC-----CAACTACCCCA	1744
Qy	810	ArgCysValSerProLeProValSerProThrSerArgIleGlnAsnProValAlaPhe	829
Db	1745	AGGAGCTGCACCACTCCCACTC-----CCACATC-----CCAAAGAACGCTGCACCCACCAACCAAGAAGC	1798
Qy	830	LeuSerSerValLeuPro-----SerLeuProAlaIleProProThrAsnAla---Met***	847
Db	1799	CTGCACCCACCGCTCCCAAGAGCTGCCCACTACCTCCCAAGGAGAGCTGCACCCACCA	1858
Qy	848	LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer	867
Db	1859	CCCCAAGAGCTCAGCCCAACCCCGAAGCTGCACCCACCACTGAGAGC	1918
Qy	868	ProGlnPro-----ValAsnAspAsnIleArgGluThrLysAsnAla	882
Db	1919	CGCACCCACCACTGCAGGAGCTGCACCCACCACTGCAGGAGCCACCACTCCCA	1978
Qy	883	ValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGlnGlu	902
Db	1979	CCCTGAGGAGCTGCTCCACCACTCCCA-----	2008
Qy	903	TyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArg	922
Db	2009	-----AGGAGCGGCTGCCAACACCCCTAAGGAGCTGCTCCCACT-2049	
Qy	923	ThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCys	942
Db	2050	ACCCCTAAGGAGCTGCTCCCACTACCTTAGGAGCTGCTCCCACTACCCCTAAGGAG	2109
Qy	943	IleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVal	962
Db	2110	ACTGCTCCAACTACCCCTAAGGAGCTGCTCCCACTACCTCAAGGAACCTGCACCACT	2169
Qy	963	ThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGly	982
Db	2170	ACT-----CCCAAGAGCTGCTCCCAAG-----	2193
Qy	983	LysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCys	1002
Db	2194	-----GAGCTGCACCCACCACTCC-2211	
Qy	1003	SerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAla	1022
Db	2212	ACCAAGAGGAGCCACCTCCCACTCTGCAAGCGCTGCCAACTACCT-2259	
Qy	1023	AsnProGlnGlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleAr	1042
Db	2260	-----CCTAAGGGAGCTGCTCCCACTACCTTACGAGCTGCTCCCACTACCT-2313	
Qy	1042	gSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgAs	1062
Db	2314	GAGCTGCTCCAA-----CTACCCCTAAGGAGCTGCTCCCACTACCTCAAGAA	2364
Qy	1062	pLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMe	1082
Db	2365	CCTGCACCACTACTCCCAAGAGCTGCTGCCCAAGGAGCTTGACCCACCACTCAAGAA	2424
Qy	1082	tValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProPr	1102
Db	2425	GGGCCACAT-----CCACC-2439	
Qy	1102	oGluLeuThrTrpLeuLeu-----As-1109	
Db	2440	ACCTTCAGACAGCTGCTCCCACTACCTTAGGAGACTGCTCCCACTACCCCAAGAG	2499
Qy	1109	nGlyInProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHi	1129
Db	2500	CCTGCACCACTACCCCAAGAGCTGCTCCCACTACTCTGAGACAC-----2548	

[illegible]

Db	1108	-----ACCATCAAGTCTGCACCCACACACC-----					1132
Qy	472	yThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysProArgSerMetal	492				
Db	1133	-----CCAAGGAGCCTGCA	1146				
Qy	492	aGluPro-GluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyc	512				
Db	1147	CCCACCACCAAGTCTGCACC-----	1169				
Qy	512	ysPheThrCysThrAlaSerAsnLysTyrglyThrValSerSerIleAlaGlnLeuHisv	532				
Db	1170	-----CACCACCT	1176				
Qy	532	alArgIyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThra	552				
Db	1177	CCCAAGGAGCCTGCACCCACCACCAAGGAGC-CTGCAC-----	1216				
Qy	552	snLeuAlaAlaIleGluProGlnProSerProProHisSerGluProProSerValGlu	572				
Db	1217	-----CCACCACCTCCCAAGGAGCCTGCACCCACCACCAAGGAGCCTGC	1262				
Qy	572	lnProProLysProLysLeuGluGlyValLeuValAsnHisAsnGluProArgSerSerS	592				
Db	1263	ACCACCAACCACCAAGTCTGCACCCACCACCTC-----CCAAGGAGC-----	1303				
Qy	592	erArgIleGlyLeuArgValHisPheAsnLeuProGluAspAspLysGlySerGluAlas	612				
Db	1304	-----CTGCACCC-ACCACCCCAAG--AAGCTGCCCCCACTACCC	1342				
Qy	612	erSerGluAlaGlyValIleThrArgGlnThrArgProAspSer***GlnGluArgp	632				
Db	1343	CAAAGGAGCCTGCACCCACCACCTCCCAAGGAGCCTACACCACCACTCCCAAGGAG-----	1398				
Qy	632	heAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProProProv	652				
Db	1399	---CCTGCACCCACCACCAAGGAGCCTGCACCCACCACCTCCC---AAAGAGCCTGCACCCA	1453				
Qy	652	alLeuAlaLysProLysLeuAspSerThrGlnLeuGlnLeuHisAsnGlnValLeu	672				
Db	1454	CTGCCCCCAAGAGCCTGCCCACTACCCCC-----	1485				
Qy	672	euGluGlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe***	692				
Db	1486	-----AAGGAGCCTGCACCCACCACCTCCCAAGGAG-----	1515				
Qy	692	etThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysA	712				
Db	1516	-----CCTGCACCCACCACCACTCC-----AAGG	1537				
Qy	712	laProSerSerGlnThrPheSerLeuAlaArgProLysTyrrPhePheProSerThrAsnt	732				
Db	1538	AGCCTTCACCCACCACT-----CCCAAGGAGCCTGCACCCACCAACCA	1582				
Qy	732	hrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer-----SerT	750				
Db	1583	AGTCTGCACCCACCACCTACCAAGGAGCCTGCACCCACCACCTACCAAGTCTGCACCCACCA	164				
Qy	750	hrProGlnThrIleGlnArgThrValSerLysGluSerLeuValSerHisProSerV	770				
Db	1643	CTCCCAAGGAGCCTTCACCCACCACCAACCAAG-----	1674				
Qy	770	alGlnThrLysSerProGlyClyLeuSerIleGlnAsnGluProGluProGlyProT	790				
Db	1675	-----GAGCCTGCACCCACCACTCCCA	1690				
Qy	790	hrGluProThr-ProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnPro	809				
Db	1697	AGGAGCCTGCACCCACCACCC---CCCAAGAGCCTGCC-----CACTACCCCA	1740				
Qy	810	ArgCysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPhe	829				

Db 1745 AGGAGCCTGCACCCA---CCACTC---CAAAGGAACCTGCACCCACCCACCCACCAAGAAGC 1798
Qy 830 LeuSerSerValLeuPro---SerLeuProAlaIleProProThrAsnAla---Met*** 847
Db 1799 CTGCACCCACCGCTCCCAAGAGCCTGCCCACTACCACTACCAAGAGGACTGCACCCACCA 1858
Qy 848 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer 867
Db 1859 CCCCCAAGAAGCTCAGCCGCCACCCAGAGCTGCACCCACCCACCCCTGAGAGC 1918
Qy 868 ProGlnPro-----valAsnAspAspAsnIleArgGluThrLysAsnAla 882
Db 1919 CCGCACCACCAACCCCTGAGGAGCTGCACCCACCCACCCCTGAGGAGCCACACCCACCA 1978
Qy 883 ValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGlu 902
Db 1979 CCGCTGAGAGCCTGCTCCACCACTCCCA----- 2008
Qy 903 TyrLysIleSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArg 922
Db 2009 -----AGCAGCGGCTCCCAACACCCCTAAGGAGCCTGCTCCAACCT- 2049
Qy 923 ThrProValaspGluSerAspGluIleGlnHisAspGluIleProThrGlyLysCys 942
Db 2050 ACCCTAAGGAGCCTGCTCCCACTACCCCTAAGGAGCCTGCTCCAACCTACCCCTAAGGAG 2109
Qy 943 IleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVal 962
Db 2110 ACTGCTCCAACCTACCCCTAAGGAGCCTGCTCCAACCTACCCCTAAGGAGCCTGCTCCAACCT 2169
Qy 963 ThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGly 982
Db 2170 ACT-----CCCAAGAGCCTGCCCCCAAG----- 2193
Qy 983 LysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCys 1002
Db 2194 -----GAGCTTGACCCACCCACC 2211
Qy 1003 SerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAla 1022
Db 2212 ACCAAGGAGCCCATCCCAACCTCTGACAAAGCCCGCTGCCAATACC----- 2259
Qy 1023 AsnProGlnGlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArg 1042
Db 2260 ---CCTAAGGGAGCTGCTCCAACCTACCCCTAAGGAGCCTGCTCCA-----CTACCCCTAAG 2313
Qy 1042 gSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgAs 1062
Db 2314 GAGCCTGCTCCA-----CTACCCCTAAGGGAGCTGCTCCAACCTACCCCTCAAGGAA 2364
Qy 1062 pLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMe 1082
Db 2365 CTGCACCCACTACTCCCAAGAGCCTGCCCAAGGAGCTTGACCCACCCACCCACCAAG 2424
Qy 1082 tValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProPr 1102
Db 2425 GGGCCACAT-----CCACC 2439
Qy 1102 oGluLeuThrTrpLeuLeu-----As 1109
Db 2440 ACCTCTGACAAGCCTGCTCCAACCTACACCTAAGGAGACTGCTCCAACCTACCCCAAGGAG 2499
Qy 1109 nGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValH1 1129
Db 2500 CTGCACCCACTACCCCAAGAGCCTGCCCAACTACTCTCTGAGACAC----- 2548
Qy 1129 sSerLeuLeuIleAspProLeu 1136
Db 2549 -----CTCCTCCAACCACTT 2563
RESULT 14
US-07-757-022B-1
; Sequence 1, Application US/07757022B

; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/757,022B
; APPLICATION NUMBER: 19910910
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5008 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4212
; US-07-757-022B-1
Alignment Scores:
Pred No.: 2,33e-10 Length: 5008
Score: 275.00 Matches: 229
Percent Similarity: 31.03% Conservative: 103
Best Local Similarity: 21.40% Mismatches: 376
Query Match: 3.99% Indels: 365
DB: 4 Gaps: 43
US-09-818-990B-2 (1-1320) x US-07-757-022B-1 (1-5008)
Qy 116 AsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGluSerProGln 133
Db 304 AGTTTCTGTGCAGAGTGTCATAATCCACATCCACCATCTTCAAGAAAGACCTCCA 363
Qy 134 GluAlaLysArgProGlnTyrCysSerGluThrGlnSerLysValPheLeuAsnLys 153
Db 364 -----CCTTCAGGAGCATCTCAAC----- 384

QY 154 AlaAlaaspPheIleGluGluLeuSerSerLeuPheLysSerHisSerLysArgIle 173
Db 385 -----ATCAATCAACAACAAAGTTCACCC 411
QY 174 ArgProArgAlaCysLysAsnHisLysSerLysLeuGluSerGlnAsn----- 189
Db 412 AAACCCACAAGAAAGAGACTAAGAAAGTATAGAAATCAGAGAAATAACAGAGAA 471
QY 190 -----LysValMetGlnGluAsnSerSerSerSerAsp----- 201
Db 472 CATTCCTTTTGAAATCAAGAGCTCCTCCTCCTCCTCTCTCTCTCTCTCTCTCA 531
QY 202 -----LeuSerGluArgGluArgSerSerValProIleProIlePro 216
Db 532 ACAATTTGGAAATCAAGTCTTCACAAATTCAGCTGCTAATAGAGAAATACAGAGAAA 591
QY 217 AlaaspThrArgAspAsnGluValAsnHisAlaLeuGluGlnGluAlaLysArgArg 236
Db 592 CTCAAAGTAAAGATACAAAGAGACAGAACTAAAGAAAGAACTACCCCAACACCA 651
QY 237 GluAlaGluGlnAlaLaserGluAlaAlaGlycylAsp-----ThrThrProGly 253
Db 652 GTTCTAGATGACCTGGAAGTGGACATGGTCACTTCAAGTCAACAACCTCTGAC 711
QY 254 SerSerProSerSerLeuTyrTyrGluGluProLeuGlyGlnProProArgPheThrGln 273
Db 712 ACGTCTACCCCAAA-----CACATAAAGTCAGCATCTCCCAAGATCACAA 762
QY 274 LysLeuArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValVal 293
Db 763 GCAAAACCAATAATCCAGACCC----- 786
QY 294 GlyIleProProGlnValArgTyrTyrCysGluGlyLysGluLeuGluAsnSerPro 313
Db 787 AGCTTCCACT----- 798
QY 314 AspileHisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGlu 333
Db 799 -----AAT 801
QY 334 GluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSer 353
Db 802 TCGATACATCTAAAGAGAGCTTTGACAGTGAATAAAGAGACAAACAGTGTAAACTAAA 861
QY 354 AlaGluIleTyrIleGluGlyValSerSerSerAsp---SerGluGlyAspProAsnLys 372
Db 862 GAACTACTACACAATAACAGACTTCACTGATGATGGAAGAGAGACTACTCTCGCT 921
QY 373 GluGluMetAsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAla 392
Db 922 AAAGAGACAAAGATATAGAGAAACATCTGCTAAAGATTAGCACCCACATCTAAAGTG 981
QY 393 ValIleProProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThr 412
Db 982 CTGGCTAAACCT---ACACCCCAAGCTGAACACTACACCAAGGCCCTGCTCTCACCACT 1038
QY 413 IleGlnGlnCysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIle 432
Db 1039 CCCAAGAG---CCACGCCCCACCACTCCCAAGGAGCTGCATCTACCAACCCCAAGAG 1095
QY 433 AlaAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuVal 452
Db 1096 CCCACACCTACC----- 1107
QY 453 ValPheGluCysArgValLysGlyAlaProSer---ProLysValGluTyrTyrArgGluGlu 472
Db 1108 -----ACCATCAAGTCTGCACCCACCACTCC----- 1132
QY 472 yThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAl 492
Db 1133 -----CCAGGAGCCTGCA 1146
QY 492 aGluPro-GluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyC 512

Db 1147 CCCACCACCAAGTCTGCAC----- 1169
QY 512 ysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisV 532
Db 1170 -----CACCACCT 1176
QY 532 alArgGlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrA 552
Db 1177 CCCAAGGAGCTGCACCCACCAAGGAGC-CTGCAC----- 1216
QY 552 snLeuAlaAlaIleGluProGlnProSerProProHisSerGluProProSerValGluG 572
Db 1217 -----CCACCACCTCCCAAGGAGCTGCACCCACCAAGGAGCCTGC 1262
QY 572 lnProProLysProLysLeuGluValValLeuValAsnHisAsnGluProArgSerSerS 592
Db 1263 ACCCACCACCAAGTCTGCACCCACCACTC-----CCAAGGAGC----- 1303
QY 592 erArgIleGlyLeuArgValHisPheAsnLeuProGluAspAspLysGlySerGluAla 612
Db 1304 -----CTGCACCC-ACCACCCCAAG---AAGCCTGCCCAACTACCC 1342
QY 612 erSerGluAlaGlyValValThrArgGlnThrArgProAspSer***GlnGluArgP 632
Db 1343 CCAAGGAGCTGCACCCACCACTCCCAAGGAGCCTACACCCACCACTCCCAAGGAG--- 1398
QY 632 heAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProProProv 652
Db 1399 --CCTGACCCACCAAGGAGCTGCACCCACCACTCC---AAAGAGCCTCACCCCA 1453
QY 652 alLeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeu 672
Db 1454 CTGCCCCAAGAAGCCTGCCCAACTACCC--- 1485
QY 672 euGluGlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe*** 692
Db 1486 -----AAGGAGCTGCACCCACCACTCCCAAGGAG----- 1515
QY 692 etThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysA 712
Db 1516 -----CCTGCACCCACCACTCCCAAGGAG-----AAGG 1537
QY 712 laProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnT 732
Db 1538 AGCCTTCCACCCACCT-----CCCAAGGAGCCTGCACCCACCACTCCCAAGGAG 1582
QY 732 hrThrAlaAlaThrValAlaProSerSerProValPheThrLeuSer-----Sert 750
Db 1583 AGTCTGCACCCACCACTACCAAGGAGCCTGCACCCACCACTTACCAAGTCTGCACCCACCA 1642
QY 750 hrProGlnThrIleGlnArgThrValSerLysGluSerLeuValSerHisProServ 770
Db 1643 CTCCTCAAGGAGCCTTACCCACCACTCCCAAGGAG----- 1674
QY 770 alGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyProT 790
Db 1675 -----GAGCCTGCACCCACCACTCCCA 1696
QY 790 hrGluProThr-ProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnPro 809
Db 1697 AGGAGCCTGCACCCACCACTCCCAAGGAGCCTGCACCCACCACTCCCAAGGAGCCTGCACCCACCA 1744
QY 810 ArgCysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPhe 829
Db 1745 AGGAGCCTGCACCA---CCACTC---CCAAGGAACCTGCACCCACCACTCCCAAGGAGC 1798
QY 830 LeuSerSerValLeuPro---SerLeuProAlaIleProProThrAsnAla---Met*** 847
Db 1799 CTGACCCACCGCTCCCAAGGAGCCTGCACCCCACTTACCCCAAGGAGACTGCACCCACCA 1858
QY 848 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer 867
Db 848 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer 867

```
Db 1859 CCCCCAAGAAGCTACGCCCCACCCACCCCGGAGAGCTCGCACCCACCACCCTGAGAACG 1918
Qy 868 ProGlnPro-----ValAsnAspAspAsnIleArgGluThrLysAsnAla 882
Db 1919 CCGCACCACACCCCTGAGAGCTCGCACCCACCACCCTGAGGAGCCACACCCACCA 1978
Qy 883 ValIleArgAspLeuGlyLysIleThrPheSerAspValArgProAsnGlnGlnGlu 902
Db 1979 CCGCTGAGGAGCTGCTCCACCACTCCA----- 2008
Qy 903 TyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArg 922
Db 2009 -----AGCAGCGGCTCCCAACACACCCCTAAGGAGCGCTCTCCAACCT- 2049
Qy 923 ThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCys 942
Db 2050 ACCCTTAGGAGCGCTCTCCAACCTACCCCTAAGGAGCTGCTCCAACCTACCCCTAAGGAG 2109
Qy 943 IleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVal 962
Db 2110 ACTGCTCCAACCTACCCCTAAAGGAGCTGCTCCAACCTACCCCTCAAGGAACCTGCACCCACT 2169
Qy 963 ThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGly 982
Db 2170 ACT-----CCCAAGAAGCTGCCCCCAAG----- 2193
Qy 983 LysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCys 1002
Db 2194 -----GAGCTTGACCCACCACCACTACC----- 2211
Qy 1003 SerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAla 1022
Db 2212 ACCAAGGAGCCACATCCACCACTCTGACAAAGCCCGCTCCAACCTACC----- 2259
Qy 1023 AsnProGlnGlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleAr 1042
Db 2260 ---CCTAAGGGAGCTGCTCCAACCTACCCCTAAGGAGCGCTGCTCCA-----CTACCCCTAAG 2313
Qy 1042 gSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgAs 1062
Db 2314 GAGCTGTCTCCA-----CTACCCCTAAGGGAGCTGCTCCAACCTACCCCTCAAGGAA 2364
Qy 1062 pLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMe 1082
Db 2365 CCGTGCACCCACTACTCCCAAGAGAGCTGCCCCCAAGAGCTTGACCCACCACCAACCAAG 2424
Qy 1082 tValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProPr 1102
Db 2425 GGGCCCACT-----CCACC 2439
Qy 1102 oGluLeuThrTrpLeuLeu-----As 1109
Db 2440 ACCTGTGACAGCGCTGCTCCAACCTACACCTAAGGAGACTGCTCCAACCTACCCCAAGGAG 2499
Qy 1109 nGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHI 1129
Db 2500 CTGCACCCACTACCCCAAGAGCGCTGCTCCAACCTACTCTCTGAGACAC----- 2548
Qy 1129 sSerLeuLeuLeuAspProLeu 1136
Db 2549 -----CTCTCTCAACCACTT 2563

RESULT 15
US-09-877-730-31
; Sequence 31, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
```

```
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 3874
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-31

Alignment Scores:
Pred. No.: 2,63e-10 Length: 3874
Score: 272.00 Matches: 241
Percent Similarity: 29.06% Conservative: 153
Best Local Similarity: 17.77% Mismatches: 433
Query Match: 3.95% Indels: 529
DB: 4 Gaps: 47

US-09-818-990b-2 (1-1320) x US-09-877-730-31 (1-3874)
Qy 229 GluGlnGlnGluAlaLysArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGly 248
Db 23 GAGCACCAGCCTCGCCCGCAGAGACGCTCGCCACGCGCGTGGAGCCGGAGCGGGGG 82
Qy 249 AspThrThr-----ProGlySerSerProSerSer 258
Db 83 AGCCAGGCTGCGTGCACGACGCGCAGAGCAGAGCGCGCGGGGGCCGAGACG 142
Qy 259 LeuTyrTyrGluGluProLeuGlyGlnProArgPheThrGlnLysLeuArg----- 276
Db 143 -----GACAGGGGCTCTGGCGCGCGGGAGCATGCCCGCGCGCTACGCTGATG 193
Qy 277 -----SerArgGluValProGluGly----- 283
Db 194 GCGCTCTCTGCGACCCCTCGCGCGGCTGGAGCGCGGGGATGCTGCTCGCGCGGCTC 253
Qy 283 ----- 283
Db 254 CTGCTCTCTGCTGMRGCTCAGTCTTGTCCAGAGTGTGGTCTTAGCGAACTGCTCTTT 313
Qy 284 -----ThrArg-----ValGlnLeuAspCysIle 291
Db 314 GTAAAGAACACAGGATGTAACTGTCAAGAAAGGACCCAGCTGTTTAGATTGCCAG 373
Qy 292 ValValGlyIleProProGlnValArgTrpTyrCysGluGlyLysGluLeuGluAsn 311
Db 374 GCTCAGCGAAGTTCCTATTAAAGTTCATGTTGAAAATGGAGCAAAATGCTGAA 433
Qy 312 SerProAspIleHisIleValGlnAlaGlyAsnLeuHisSerLeuThrIle-----Ala 329
Db 434 AATAACGATCGAGGTCTTTCTTAACGGCTCTTTATATACATCAGTGGAGTGAAGGCAGG 493
Qy 330 GluAlaPheGluGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThr 349
Db 494 CGAGGAGACGATCCGATGAAGGATTTATCAGTGTGTCGATTTGCAATGAACAAATGGAGCC 553
Qy 350 Asp-----SerThrSerAlaGluIleTyrIleGluGly 360
Db 554 ATTCTTACTCAAAAGCTCATCTTGCCTTATCAACTATTCTGCAATTGAAAGTCCAGCCA 613
Qy 361 ValSerSerSerAspSerGluGlyAspProAsnLys----- 372
Db 614 ATTTCCACTGAGTCCAGAGGTGGAGTTCCTCGATTTGTCATCAAGATTTTCATCCAC 673
Qy 373 -----GluGluMetAsnArgIleGlnLysPro-----AsnGlu 383
Db 674 CCTCTCGAGTCATACATGGAGTTCAATCGGACAACTCTACCTATGACTATGAGCAGG 733
```

```
QY 384 ValSerSerProThr----- 389
Db 734 ATRACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTACAGCAAAAGGATTCTGGA 793
QY 390 -----ThrSer 391
Db 794 AATTATCGTTGTATTGCTGCCACTGTAGCCACCAGCGTAGTAAAGTATGGAGGCTCGCTA 853
QY 392 AlaValIleProAlaValProGlnAlaGlnHis-----LeuValAlaGlnPro 408
Db 854 ACTGTGATTCACGCTAAGGAGTCAAAATCTCTCCACACACACARCAATATTAGCAGGTCCA 913
QY 409 ArgValAlaThr----- 412
Db 914 CAGAACATAACAACATCTCTTCATCAGACTGTAGTTTGGAAATGCATGGCCACAGGAAT 973
QY 412 ----- 412
Db 974 CCCAAACCAATCATTTCTTGGAGCGCCTTGGATCACAATCCATTGATGCTTTAATACT 1033
QY 413 -----Ile 413
Db 1034 CGSGTACTTGGAAATGTAATCTCATGATATCTGATCTGATCTCAGGCTACAAACATGCTGGAGTA 1093
QY 414 GlnGlnCysGlnSerProThr-----AsnTyrLeuGlnGlyLeuAspGlyLys 429
Db 1094 TATGTTTGTGGGCCACTACCCCTGGCCACACGCAACTTTACAGTTGCTATGGCAACTTTA 1153
QY 430 ProIleIleAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGly 449
Db 1154 ACTGTATTAGTCTCTCTCTTATTGTTGAATGGCCAGAAAGTTTAAACAGGCCCTCGAGCT 1213
QY 450 GlnLeuValValPheGluCysArgValLysGlyAlaProSerProLysValGluTyr 469
Db 1214 GGCACCTCGATTGTGTGTGTCAGGAGGAAGTCCCTCTCCCAAGATGTCATGGTTG 1273
QY 470 ArgGluGlyThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArg 489
Db 1274 AAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAATAATGTACACAGTAAA----- 1327
QY 490 SerMetAlaGluProGluGluIleCysThrLeuValIleAlaGluValPheAlaGluAsp 509
Db 1328 -----TTGGTAATTAAACCAGATTAATTCCTGGAAGT 1357
QY 510 SerGlyCysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGln 529
Db 1358 GATGCTATTATCAGTGCATGGCTGAGAAATAGCCAGAGATCTATTATCTAGAGCCAGA 1417
QY 530 LeuHisValArgGlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSer 549
Db 1418 CTGACTGTAGTGTGCAGAAAGC----- 1441
QY 550 ThrThrAsnLeuAlaAlaIleGluProGlnProSerProPro-----HisSerGlu 566
Db 1442 -----AGACCCAGTCTCCCTATAATGATACATGCTGAA 1474
QY 567 ProProSerValGluGlnProProLysProLysLeuGluGlyValLeuValAsnHisAsn 586
Db 1475 ACATGTC-----AGCTCAGCAATCTTTTAGCCTGGGAG 1510
QY 587 GluPro-----ArgSerSerArgIleGlyLeuArgValHisPheAsnLeuProGlu 604
Db 1511 AGGCCACTTTTAAATTCAGACAAAGTCAATTCCTTCTGTACACTACATCAAGCAGAA 1570
QY 605 AspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrArgGlnThrArg 624
Db 1571 GGTTTAAATAATGAAGAGTATCAAGTAGTCATCGAAATGACACAACTCATATATTATT 1630
QY 625 ProAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGluProSerPhe 644
Db 1631 GATGAC----- 1636
QY 645 ProValLysGluProProPro-----ValLeuAlaLysProLysLeuAsp 659
```

```
Db 1637 -----TTAGAGCTGCCAGCAATTATACTTTCTACATATTAGCATATATGCCAATGGGA 1690
QY 660 SerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGlnHisGlnLeuGlnAsn 679
Db 1691 GCCAGCCAGATGCTGTGACCATGTGACACAGATACTCTAGAGGATGTTCCCTCGAGA--- 1747
QY 680 ProProSerSerProLysGluPheProPhe***MetThrValLeuAsnSerAsnAla 699
Db 1748 CTCCTCGAAATTAGT-----TTGACAAGTTCGAAGT 1777
QY 700 ProProAlaValThrThrSer***LysGlnValLysAlaPro----- 713
Db 1778 CCCACTGATATTCTCATCTCTCGCTGCCAATCCAGCCAAATATCGCGGGGCGCAAGTG 1837
QY 714 -----SerSerGlnThrPheSerLeu 720
Db 1838 GTGCTGATCGCTTGTCTTTCGCGCTTAAGTACTGAGAAATCCAATCCAAAGTCTGGAGCTC 1897
QY 721 -----AlaArgProLysTyrPhePheProSer 729
Db 1898 CCGGGGACCACGATGAGTACCTTTTGGAGGCTGAACCTGCACAGTGTCTACCTGGTT 1957
QY 730 ThrAsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSerSer 749
Db 1958 CGGATTACTGCTGCCACAGAGTGGGCTGGAGAGTCATCAGTATGGACTTTCACATAGG 2017
QY 750 ThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSer 769
Db 2018 ACGCCCAAA-----GCT 2029
QY 770 ValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyPro 789
Db 2030 ACAGCTGTAAGCCCTTAAGTCTCCAGAGTTGCAATTTGGAGCCTCG----- 2077
QY 790 ThrGluProThrProProPheThrPheSerIleProSerGlyAsnGlnPheGlnPro 809
Db 2077 ----- 2077
QY 810 ArgCysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPhe 829
Db 2078 AACTGTACCACC---ATTCTGTGAGTGGCAGCAAGATGTAGAGGACACAGCTGCTATT 2134
QY 830 LeuSerSerValLeuProSerLeuProAlaIleProProThrAsnAlaMet***LeuPro 849
Db 2134 ----- 2134
QY 850 ArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGln 869
Db 2135 -----CAGGGCTACAAGCTGTACTACAAGGAAGAGGCGAG 2170
QY 870 ProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLys 889
Db 2171 CAGAGAATGGGCCATTTTCTGGATACCAAGGACCTACTC----- 2212
QY 890 LysIleThrPheSerAspValArgProAsnGlnGlnGluTyrLysIleSerSerPheGlu 909
Db 2213 -----TATACCTCAGTGGCTTAGAC 2233
QY 910 GlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAsp 929
Db 2234 CCCAGAAGA-----AATATCATGTGAGACTCTCGCTTACAACACATAGAC 2281
QY 930 AspGluIleGlnHisAspGlu---IleProThrGlyCysCysIleAlaProIlePheAsp 948
Db 2282 GATGGCTATCAGGAGATCAGACTGTGTCAGCACTCCAGATGC----- 2323
QY 949 LysArgLeuLysHisPheArgValThrGluGlySerProValThrPheThrCysLysIle 968
Db 2324 -----GTGCTGTCTGTCGATCGCATG 2344
QY 969 ValGlyIleProValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArg 988
```

```
Db 2345 GTCCCTCTCTCCACCACACCC----- 2365
Qy 989 AsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGlu--- 1007
Db 2366 ---CACCATCTCTATCGCAAGGCTAACACCTTCATCTTCCATCTTCCGTGCATGGAGGAGG 2422
Qy 1008 ---SerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAspProGlnGly 1026
Db 2423 CCTGCATTCACCGCTGCACAAATCATTAATACACCATC---CGGTGTAATCTCTGTGGC 2479
Qy 1027 ArgIleSerCysSer-----GlyHisLeuMetValGln 1037
Db 2480 CTGCAGAATGCTTCTTGGTCTCTACCTTCAACATCAGAACTCACATGTTGTTCAA 2539
Qy 1038 SerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArg 1057
Db 2540 GGTCTAGNACCAACACCAATAGCAATTTGCCGTTGCATTACAT----- 2584
Qy 1058 ValGlnGluArgAspLysGluProLeuGlnGluArgPheArgProHisPheLeuGln 1077
Db 2585 ---GTGGATCAGCTTTCAGTCTTGGAGCCCTGTAGTCTACCATTTACTCTTCCAGAA 2641
Qy 1078 AlaPro---GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysVal 1096
Db 2642 GCACGACGAGCCACCCAGTTGGAGTAAAGTGACATTAATAGAGGATGACACACTGCCCTG 2701
Qy 1097 SerGlyLeuProProGlu-----LeuThrTrpLeuLeu 1108
Db 2702 GTTCTTGGAAACCCCTGTAGTGGCCAGAACAGTTGTGACCCGCTATATCTTATAT 2761
Qy 1109 AsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyVal 1128
Db 2762 GCATCTAGGAAGGCTGGATTGCGAGGAGTGGCAGGTCTTACACCGTGAAGGGCAATA 2821
Qy 1129 HisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAla 1148
Db 2822 ACCATGGCTTTGCTAGAAAACCTTGGTAGCAGGAATGTGTACATTTGCAAGATATCTGCA 2881
Qy 1149 ThrAsnLysThrGlyGlnAsnSerPhe-----SerLeuGluLeuSerValValAlaLys 1166
Db 2882 TCCAAATGAGTGGAGAGGAGGCCCTTTTCAAAATCTGTGGAGCTGGCAGTACTTCCAAAG 2941
Qy 1167 GluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys----- 1181
Db 2942 GAAACCTCTGAATCAAAATCAGAGGCCCAAGGCTTAGATTCTCGCTGATGCCAAAGTTAT 3001
Qy 1182 -----GlyValProGluGlyHisProVal 1189
Db 3002 TCAGGATATTACCATCTGGACCAAAATCAATGACTGCGCATGTGTAGGTGTTGGCATA 3061
Qy 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsn 1209
Db 3062 GCCTTGACCTGCATCTCATC----- 3082
Qy 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAla 1229
Db 3082 ----- 3082
Qy 1230 CysLeuLeuIle-----GlnProAlaLysLysSerAspAlaGlyTrpTyrThr 1245
Db 3083 TGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAATCATCTGCT-----TCC 3133
Qy 1246 LeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAla 1265
Db 3134 AAGACGGCAGAGATGGAAT----- 3154
Qy 1266 GlnTrpHisHisGlnIleProProMetSerValArgProSerGlySerArgTyrGly 1285
Db 3155 -----CAACAGTTACCTCGTACCAGTGCCTCTTAGCTAGTAGGAAATGAGTAGGA 3205
Qy 1286 SerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerMetGluSerThrMet 1305
Db 3206 -----AAGAACCTGGAA-----GGAGCTGTAGGAATGAAGATCTTTAATG 3247
```

Qy 1306 ValTyr-SerCysSerSerArgSerValValGluSerAspGluLeu 1320
Db 3248 CCAATGATCATGCCAAACAGCTTCATTCATGCAAGGAGGAAGCTG 3293

Search completed: November 30, 2002, 21:35:34
Job time : 178 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 30, 2002, 20:59:00 : Search time 93 Seconds
(without alignments)
5466.155 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence: 1 MQDDSTEAISTISQLRESY.....MESTMYSCSRVSDESL 1320

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09818990/runat_26112002_093406_23337/app_query.fasta_1.1479
-DB=Published Applications_NA -QFAST=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum62
-TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09818990.scgn.1.1.47 @runat_26112002_093406_23337
-NCPU=6 -ICPU=3 -NO_XLPAY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1510.5	21.9	1823	10	US-09-925-299-231 Sequence 231, App
2	1394	20.2	2059	10	US-09-764-853-218 Sequence 218, App
3	1133.5	16.5	702	10	US-09-764-853-397 Sequence 397, App
4	557.5	8.1	5926	10	US-09-969-708-302 Sequence 302, App

5	557.5	8.1	5926	10	US-09-954-456-522 Sequence 522, App
6	527.5	8.1	5926	10	US-09-880-107-3371 Sequence 3371, App
7	523.5	7.6	1833	10	US-09-880-192-32 Sequence 32, Appl
8	508.5	7.4	23907	9	US-10-077-130-6 Sequence 6, Appl
9	508.5	7.4	24120	9	US-10-077-130-4 Sequence 4, Appl
10	432.5	6.3	3192	10	US-09-925-300-502 Sequence 502, App
11	416.5	6.0	8883	10	US-09-905-129-1 Sequence 1, Appl
12	416.5	6.0	8883	10	US-09-905-129-5 Sequence 5, Appl
13	416.5	6.0	8883	10	US-09-905-129-7 Sequence 7, Appl
14	416.5	6.0	8883	10	US-09-991-630-1 Sequence 1, Appl
15	416.5	6.0	8883	10	US-09-991-630-5 Sequence 5, Appl
16	416.5	6.0	8883	10	US-09-991-630-7 Sequence 7, Appl
17	408.5	5.9	8180	10	US-09-905-129-8 Sequence 8, Appl
18	408.5	5.9	8180	10	US-09-991-630-8 Sequence 8, Appl
19	408.5	5.9	8262	10	US-09-905-129-6 Sequence 6, Appl
20	408.5	5.9	8262	10	US-09-991-630-6 Sequence 6, Appl
21	408	5.9	7770	10	US-09-991-630-23 Sequence 23, Appl
22	397	5.8	355	10	US-09-563-817-656 Sequence 656, App
23	396	5.7	9645	10	US-09-905-129-22 Sequence 22, Appl
24	396	5.7	9645	10	US-09-991-630-22 Sequence 22, Appl
25	390	5.7	443	10	US-09-960-352-13726 Sequence 13726, A
26	378	5.5	8106	9	US-10-077-130-1 Sequence 1, Appl
27	369	5.4	7893	9	US-10-077-130-3 Sequence 3, Appl
28	346	5.0	453	10	US-09-960-352-20 Sequence 20, Appl
29	337	4.9	431	10	US-09-960-352-8748 Sequence 8748, Ap
30	334	4.8	456	10	US-09-960-352-13256 Sequence 13256, A
31	327	4.7	2793	9	US-10-160-865-13 Sequence 13, Appl
32	326	4.7	2614	9	US-10-160-865-15 Sequence 15, Appl
33	323.5	4.7	4291	12	US-10-105-934-4 Sequence 4, Appl
34	322.5	4.7	468	10	US-09-960-352-932 Sequence 932, App
35	312.5	4.5	2259	12	US-10-105-934-6 Sequence 6, Appl
36	301	4.4	377	10	US-09-960-352-8061 Sequence 8061, Ap
37	296	4.3	407	10	US-09-983-965-508 Sequence 508, App
38	290	4.2	1083	10	US-09-764-853-373 Sequence 373, App
39	289	4.2	941	10	US-09-764-853-184 Sequence 184, App
40	285.5	4.1	417	10	US-09-960-352-933 Sequence 933, App
41	283	4.1	4682	10	US-09-783-066-8 Sequence 8, Appl
42	278.5	4.0	4978	12	US-10-044-090-89 Sequence 89, Appl
43	278	4.0	4092	12	US-10-124-557-51 Sequence 51, Appl
44	276.5	4.0	5510	10	US-09-880-107-1758 Sequence 1758, Ap
45	275	4.0	3420	12	US-10-124-557-103 Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-925-299-231
; Sequence 231, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 231
; LENGTH: 1823
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (82)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1593)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature

; LOCATION: (1714)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-925-299-231

Alignment Scores:

Pred. No.:	1.29e-93	Length:	1823
Score:	1510.50	Matches:	293
Percent Similarity:	69.09%	Conservative:	87
Best Local Similarity:	53.27%	Mismatches:	118
Query Match:	21.92%	Indels:	52
DB:	10	Gaps:	7

US-09-818-990B-2 (1-1320) x US-09-925-299-231 (1-1823)

Qy	818	SerProThrSerArgIle-----GlnAsnProValAlaPheLeuSerSerValLeu	834
Db	2	TCGCCCTGCCACCCGCTTCGGCCACGCCACGCCCGCGCGCTCCTCAGCGCTCTGGT	61
Qy	835	ProSerLeuProAlaIleProProThrAsnAlaMet**LeuProArgSerAlaProSer	854
Db	62	CCCTCGCAGCCG-----CGCGCG-----GNCGGGTCAACGCCCTGGGGCTGCCAAG	109
Qy	855	MetProSerGlnGlyLeuAla-LysLysAsnThrLysSerProGlnProValAsnAspAs	874
Db	110	GTGTACACCCCGCAGATTTCCAAAGAGGCCAGTAGTAAGTCTAGAAATAGCCTCGATGA	169
Qy	874	pAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSe	894
Db	170	GGAATTCAGGCACAAAGATGCTGTTATTCAAGACCTGGAAACGAACACTTCGCTTCAA	220
Qy	894	rAspValArgProAsn-----	899
Db	230	GGAGGACCTCCTGAACAATGGCCAGCGGAGGTTTAACATACGAAGAAGAAATGGCTCGTCG	289
Qy	899	-----	899
Db	290	ACTGCTAGGTGCTGACAGTGCAACTGCTTTTAATATTACGAGCCAGAGGAACACAGC	349
Qy	900	-GlnGlnGluThrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheAr	919
Db	350	TAATCAGGAATACAAAGTCTCCAGCTGTCAACAGAGACTCATCAGTGAATAGACTACAG	409
Qy	919	gLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProTh	939
Db	410	GCTAGAAAGGCTCTCTGTGGATGAATCAGGTCATGAAGTTCAGTAGGAGTGTCGCTGT	469
Qy	939	rGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluCl	959
Db	470	GGAATGGNAATGGCACCATCTCTTGAGATGAAGCTGAACATTCACAGATCTTTGAGGG	529
Qy	959	ySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrrPpH	979
Db	530	AATGCCAGTAACCTTCACATGTAGAGTGCTGGAATCCAAAGCCAAAGATCTATTGGTT	589
Qy	979	eLysAspGlyLysGlnIleSerLysArgAsnGlnHisCysLysMetArgArgGluCluYAs	999
Db	590	TAAAGATGGGAACGATCTCTCCAAGAGTGTATCACTACACATTCAAAGAGATCTCGA	649
Qy	999	pGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrrIle	1019
Db	650	TGGGACCTGCTCCCTCCATACCACAGCCCTCCACCTAGATGATGTTGGAATATACAAT	709
Qy	1019	eMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLe	1039
Db	710	TATGGCTGCAAAACCCCTCAGGGCCGATCATAGTTGTACTGGACGGCTTAATGGTACAGGCTGT	769
Qy	1039	uProIleArgSerArgLeuThrSerAlaGlyGlnSerHis-----Ar	1053
Db	770	CAACCAAGAGGTGGAAGTCCCGGTCTCCCTCAGGCCATCTCTCATGTCAGNAGGCCTCG	829
Qy	1053	gGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPr	1073
Db	830	TTCTAGATCAAGGGACAGCTGGAGACGAAATGAACCAATTCAGGACGATTTTCAGACC	889

Qy	1073	oHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAs	1093
Db	890	TCACCTTCTGCAGGCTCCTGGAGATCTGACTGTTCAAGAAGGAAAACTCTCAGAATGGA	949
Qy	1093	pCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVa	1113
Db	950	CTGCAAAAGTCAGTGGGTACCAACCCACGATCTAAGCTGGCAACTAGATGGAAGCCCGT	1009
Qy	1113	lLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIl	1133
Db	1010	ACGCCCTACAGTGCCTCAAGAATGCTGGTGGTGAGAACGGGGTGCACTCTCTGATCAT	1069
Qy	1133	eAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGl	1153
Db	1070	AGAGCCAGTCACGTCACGTGATCCGGCATCTACACATGTATAGCTACCAACCGAGCAGG	1129
Qy	1153	yGlnAsnSerPheSerLeuGluLeuSerValValAlaLysLysGluValLysLysAlaProVa	1173
Db	1130	ACAGAACATCATTCAGCGTGGAGCTTGTGGTGTCTGCTAAAGAAGACACAAACCCCGCTGT	1189
Qy	1173	lIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCy	1193
Db	1190	GTTTATTTCAGAGAGCTCCAAACACACAGAGGTGCTGATGGGTACCCAGTCCGGCTGGAAATG	1249
Qy	1193	sArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrIlePr	1213
Db	1250	TCGTGTATTGGGAGTGCACCACTCAGATATTTTGGAGAAAGAAAAATGAATCACTCAC	1309
Qy	1213	oCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIl	1233
Db	1310	TCACAGACACTGACCGAGTGAGCATGCACAGGACCAACCCAGCTACATCTGCCTGCTCAT	1369
Qy	1233	eGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGl	1253
Db	1370	TCAGGGAGCCACAAAGAAGATGCTGGGTGGTATCTATCTGTCAGCCCAAGATGAAGCAGG	1429
Qy	1253	yIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis-----GlnIl	1271
Db	1430	GATTGTGCTGTACTGCCAGGCTGCACGTTTACACCCAGTGGCATCAGCAGTCACAGAG	1489
Qy	1271	eProProMetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysGl	1291
Db	1490	CACCAAGCCAAAAAGTAGCGGCCCTCAGCCAGTCGCTATGCAGACACTTTCGGACCAAGG	1549
Qy	1291	yLeuAspIlePheSerAlaPheSerMetGluSerThrMetValTyrSerCysSerSe	1311
Db	1550	ACTAGACATCAAGACGCGTTCCAACT---GAGGCCAACCCATCTCNACCTGCACACTGAA	1606
Qy	1311	rArgSerValValGluSerAspGluLeu	1320
Db	1607	TACTGCTCTGGTAGAAAGTGAGACACCTG	1634

RESULT 2

```

US-09-764-853-218
; Sequence 218, Application US/09764853
; Patent No. US2002090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P3206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 218
; LENGTH: 2059
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-218
Alignment Scores:

```

Pred. No.: 1.1e-85 Length: 2059
Score: 1394.00 Matches: 261
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.62% Mismatches: 0
Query Match: 20.23% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-2 (1-1320) x US-09-764-853-218 (1-2059)

```
QY 1059 GlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAla 1078
Db 7 AGAAGAGAGACAAAGAGCCCTACAGAACGCTTTTCCGACCACATTTCTCCAGGCT 66
QY 1079 ProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGly 1098
Db 67 CCTGGGATATGTAGTCTATGAGGGCGCTCTGCGCTGGACTGTAGGTAGTGGT 126
QY 1099 LeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSer 1118
Db 127 TTACCGCCCGGAGCTGACATGCTACTCAATGGCCAACTGTGCTACCAGATGCCCTC 186
QY 1119 HisLysMetLeuValArgGluThrGlyValHisSerLeuLeuLeaspProLeuThrGln 1138
Db 187 CACAAGATGCTGTACGGGAGACCGGAGTCCACTCTGTCTCATTTGACCACCTCACTCAG 246
QY 1139 ArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSer 1158
Db 247 CGGACGACGAGGACCTATAAGTGCATCGCTACCAACAAACCGGGCAGAAATCTTTAGT 306
QY 1159 LeuGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeu 1178
Db 307 CTGAGCTCTCTGTAGTACAAAGAGGTGAAGAAAGCACCTGTGTATCTGGAGAACTA 366
QY 1179 GlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMet 1198
Db 367 CAGAACTGCCGGTGTCCCGAAGGCCACCCCGTGAGACTGGAGTGCCTGATAGGCATG 426
QY 1199 ProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArg 1218
Db 427 CCCCCACCTGTCTTCTACTGGAGAAAGACAATGAGACCATCTCTTCATTCACCCAGAGAGG 486
QY 1219 IleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLys 1238
Db 487 ATCAGTATGCACAGGACACAAAGGTATGCTGCTGCTTCTCATTCACCCAGCAAGAAA 546
QY 1239 SerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThr 1258
Db 547 TCAGACGCTGGATGGTACAGCTTGTACGCCAAGAAATGAAGCCGCAATCGTGTGTCGCACT 606
QY 1259 AlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIleProProMetSerValArg 1278
Db 607 GCACGGCTGGATATATACGCTCAGTGGCACCATCATGATCCACCGCCCATGTCTGTCCGG 666
QY 1279 ProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPhe 1298
Db 667 CCCAGTGGCAGTCGCTACGATCTCTCACAGTAAAGGACTTGACATATTTCTGCCTTT 726
QY 1299 SerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAsp 1318
Db 727 TCCTCCATGGAAAGCACGATGGTATTCATGCTCTTCTCGGAGTGTAGTGGAGAGTAT 786
QY 1319 GluLeu 1320
Db 787 GAACCT 792
```

RESULT 3

US-09-764-853-397
; Sequence 397, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206

; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 397
; LENGTH: 702
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (587)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (661)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (670)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-853-397

Alignment Scores:
Pred. No.: 1.14e-68 Length: 702
Score: 1133.50 Matches: 226
Percent Similarity: 96.17% Conservative: 0
Best Local Similarity: 96.17% Mismatches: 9
Query Match: 16.45% Indels: 4
DB: 10 Gaps: 0

US-09-818-990B-2 (1-1320) x US-09-764-853-397 (1-702)

```
QY 1060 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 1079
Db 3 GAAAGACAAAGAGCCCTACAGAACGCTTTTCCGACCACATTTCTTCGAGGCTCT 62
QY 1080 GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeu 1099
Db 63 GGGGATATGTAGTCTATGAGGGCGCTCTGCGGCTGGACTGTAAGGTAGTGGTTTA 122
QY 1100 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 1119
Db 123 CG-CCCCGGAGTGTAGTGTACTCAATGGCAACCTGTGTACAGATGCCCTCCAC 181
QY 1120 LysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArg 1139
Db 182 AGATGCTGTGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTACGCG 241
QY 1140 AspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 1159
Db 242 GACGA-GGGACCTATAAGTGCATCGCTACCAACAAACCGGGCAGAATCTTTTAGTCTG 300
QY 1160 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 1179
Db 301 GAGCTCTCTGTAGTACAAAGAGGTGAAGAAAGACCTGTGTATCTCTGGAGAAACTACAG 360
QY 1180 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 1199
Db 361 AACTGCCGTGTTCGGAAGGCCACCCCGTGAGACTGGAGTGCCTGATAGGATGCC 420
QY 1200 ProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle 1219
Db 421 CCACCTGTCTTACTTGAAGAAAGACAATGAGACCATCCCTTCACCAGAGAGGATC 480
QY 1220 SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSer 1239
Db 481 AGTATGCACGAGGACACAAAGGATGCTGCTGCTTCTCATTCAGCCAGCAAGAAATCA 540
QY 1240 AspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 1259
Db 541 GACGCTGGATGGTACAGTGTGTACGCCAAGAAATGAAGCCGCGCATCGNGTCTGTGACGTGCC 600
QY 1260 ArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProMetSerValArgPro 1279
```

Db 601 AGGCTGGATATATACGCTCAGTGGCACCACATCAGATCCACGCCCATGTC--TGCGGGGCC 658

Qy 1280 SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIle 1294
|||||
Db 659 AGNGCCAGTCGNTAGGATCTCTACACT-AAAGGACTTGACATA 702

RESULT 4
US-09-969-708-302
; Sequence 302, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 302
; LENGTH: 5926
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-302

Alignment Scores:
Pred. No.: 1,23e-28 Length: 5926
Score: 557.50 Matches: 265
Percent Similarity: 33.71% Conservative: 124
Best Local Similarity: 22.96% Mismatches: 435
Query Match: 8.09% Indels: 330
DB: 10 Gaps: 42

US-09-818-990B-2 (1-1320) x US-09-969-708-302 (1-5926)

Qy 195 AsnSerSerPheSerAspLeuSerGluArgGluArgSerSerValProIlePro 214
:::|::|
Db 18 AGTTGACAGCAACTTCAGGAGCTTCCACGCGAGAGCTTCAGAGCGCTTTCTGTCGCCA 77
Qy 215 -----IleProAlaAspThrArgAspAsnGluValAsnHisAlaLeuGluGln 231
:::|::|
Db 78 CTGCGCCAGTTGCCACAACAACACAGAGAGAGGCTGACCATGGG----- 125
Qy 232 GluAlaLysArgGluAlaGluGlnAlaAlaSerGluAlaLaglyGlyAspThrThr 251
:::|::|
Db 126 -----GATGTGAAGTGGTTCCTCGCTGCACACATT--TCCAAAACCTCC 167
Qy 252 ProGlySerSerProSerSerLeuTyrTyrGluProLeuGluGlnProProArgPhe 271
|||||
Db 168 CTCAGTGTGGATCCTCAAGAGTT---GACTCCATGCCCTGCACAGAGGCCCTGCTTC 224
Qy 272 ThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIle 291
|||||
Db 225 ATTTTGGCCCCCGAACCTCTGCATCAAGAGAGGCCACCGCAAGTTCGAAGGGCGG 284
Qy 292 ValValGlyIleProProGlnValArgTrpTyrCysGluGlyLysGluLeuGluAsn 311
|||
Db 285 GTCCGGGTTTACCAGAGCCCCAGTGACATGGCACAGAAAACGGCAACCATC---ACC 341
Qy 312 SerProAspIleHisIleValGlnAlaGly-----AsnLeuHisSerLeuThrIleAla 329
|||
Db 342 AGCGGGGCCGCTTCCTGCTGGATGGCGCATCCGGGGAGCTTTTCAGCCTGTGATTTCAT 401
Qy 330 GluAlaPheGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThr 349
|||||
Db 402 GCTGTCCATGAGGAGGACAGGGGAAGATATACCTGTGNAGGCCCAACCATGGCAGTGGTGCT 461

Qy 350 AspSerThrSerAlaGluIleTyrIleGluGlyValSerSerSerAspSerSerGluGlyAsp 369
:::|::|
Db 462 CGCCAGGTGACAGTGGAGTTCACAGTAGAAGGA---AGTTTGGCAAGCAGCAGTGGTCAG 518
Qy 370 Pro-----AsnLysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerPro 387
|||
Db 519 CCTGTTGTTCCAAAACCTTAGGGGATAGATT----- 551
Qy 388 ProThrThrSerAlaValIleProProAlaValProGlnAlaGlnHisLeuValAlaGln 407
:::|::|
Db 552 ---TCAGCTTCAGCAGTG-----GAG 569
Qy 408 ProArgValAlaThrIleGlnGlnCysGlnSerProThrAsnTyrLeuGlnGlyLeuAsp 427
|||
Db 570 ACCGTCCTAGCATCTGGGGGAGTGC----- 596
Qy 428 GlyLysProIleIleAlaAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSer 447
|||
Db 597 -----CCACCAAGTTTGCTACCAAGCTGGGCCGAGTTGTGGTCAAA 638
Qy 448 GluGlyGlnLeuValPheGluCysArgValLysGlyAlaProSerProLysValGlu 467
|||||
Db 639 GAAGCAGATGGGACGATCTCTGCAAGATCACTGGCGGCCCAACCCGACGTCACC 698
Qy 468 TrpTyrArgGluGlyThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLys 487
|||
Db 699 TGGCTCAAGGGAATGTTCCACTGAG-----CCGAGTGCCCGTGTG----- 740
Qy 488 ProArgSerMetAlaGluProGluGluIleCysThrLeuValIleAlaGluValPheAla 507
|||||
Db 741 -----TCTGTGTCTGAGAAGACGCGATCGAGTTTCTGGAATCCATGGAGTCAACCAA 794
Qy 508 GluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIle 527
|||
Db 795 GATCAGCTGGAGGTGTACACGTGCTGCTGGTGAACGGTCCGGGGAAGGCCCTCATGTCA 854
Qy 528 AlaGlnLeuHisValArgGly----- 534
|||
Db 855 GCTGAACCTTTCATCCAAAGGTTTGGACAGTGCCTAATAGGTGTCAGAGAGAACAAA 914
Qy 535 -----AsnGluAspLeu-----SerAsnAsn 541
|||
Db 915 GCCACCAATTTCAGATGTCCGAAAAGAGGTGACCAATGTAATCTCAAAGGAGTCCGAAGCTG 974
Qy 542 GlySerLeuHisSerAlaAsnSerThrAsnLeuAlaIleGluProGlnProSer 561
|||||
Db 975 GACAGTCTGGAGGTGAGCCCAAAAGCAAGAACTGCTCCAGCCCCCAGAGAGGTGGCTCC 1034
Qy 562 ProPro-----HisSerGluProProSerValGluGlnProProLys---ProLys 577
|||||
Db 1035 CCACCTCTGGGCTGCAACACAGCCAGCT-----CAGCCCCCAAGGAGTCCCAAG 1082
Qy 578 LeuGluGlyValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArg 597
|||||
Db 1083 CTGGAGTCAATGC-----AAGGACTCGCCCAAGACGGCC----- 1115
Qy 598 ValHisPheAsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyVal 617
|||
Db 1116 -----CCGACACCCCGTCTCTCAGAAGACTTCACGCTCC----- 1151
Qy 618 ValThrThrArgGlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThr 637
:::|::|
Db 1152 ATCACCCTGAGGCGCGCAAGAGATTTCAG----- 1178
Qy 638 LysThrProGluProSerPhePro-----ValLysGluProProProValLeuAla 654
|||||
Db 1179 -----CCGGAACCAAGACGACGCGCTGGGGTCTCTATCACCT----- 1217
Qy 655 LysProLysLeuAspSerThrGlnLeuGlnLeuHisAsnGlnValLeuLeuGluGln 674
|||
Db 1218 -----TCTGGA 1223
Qy 675 HisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe***MetThrVal 694


```
Db 1224 GAAGAGAGGAGGAGCGAGCTCCTCCCGTCCAGCCACCTTCCCCACGCGAGCGTGGC 1283
QY 695 LeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysAlaProSer 714
Db 1284 CTGGGAGCCAA-----GATGTTGTGAGCAAGGCTGCTAACAGGAAATCCCCCATG 1334
QY 715 SerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnThrThrAla 734
Db 1335 GAGGCCAGAGGATTCAGCATTCCTCCCAATTT----- 1367
QY 735 AlaThrValAlaProSerSerSerProValPheThrLeuSerSerThrProGlnThrIle 754
Db 1368 -----AGCCAGCAAGCCCA----- 1382
QY 755 GlnArgThrValSerLysGlu-SerLeuLeuValSerHisProSerValGlnThrLysSe 774
Db 1383 -----AGCCAGGAGGTCAGGAAATCAAACTCTCAAGTTCAGATGTGAAGTT 1430
QY 774 rProGlyGlyLeuSerIleGln-AsnGluProLeuProGly-----ProThrGluP 792
Db 1431 TCGGGATTCCAAGCCTGAAGTGGC-----CTGGTTCTTGAAGCACCCCGGTGAGG 1484
QY 792 roThrProProPheThrPheSerIleProSerGlyAsnGlnPhe-----GlnProA 810
Db 1485 AGACA-----GGAAGGCAGCATTTAGGTTTATGAA 1514
QY 810 rGcysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheL 830
Db 1515 GATGCTGGCTCCCATTAACCT----- 1534
QY 830 euSerSerValLeuProSerLeuProAlaIleProPro-ThrAsnAlaMet***LeuPro 849
Db 1535 -----CTGCTGCTGAAGCCCGGACC----- 1556
QY 850 ArgSerAlaProSerMetProSerGlnGlyLeuAlaLysAsnThrLysSerProGln 869
Db 1556 ----- 1556
QY 870 ProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLys 889
Db 1557 -----AGGACAGTGGG----- 1568
QY 890 LysIleThrPheSerAspValArgProAsnGlnGlnGluThrLysIleSerSerPheGlu 909
Db 1569 -----ACATACAGCTGCATGCTTCCACGCCCAAGCCAGGCTGCTGTAGTGG--- 1619
QY 910 GlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAsp 929
Db 1620 -----ACCTTCCAAGTGGAAAGGCTTGCCGTGATGGAG----- 1652
QY 930 AspGluIleGlnHisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLys 949
Db 1653 -----GTGGCCCTCTCTTCCAGT 1673
QY 950 ArgLeuLysHisPheArgValThrGluGlySerProValThrPheThrCysLysIleVal 969
Db 1674 GTCTGAGGACTCGCTGTTATGAGGCGCAGGATTTGTGCTGCAGTCTCGGTACGG 1733
QY 970 GlyIleProValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsn 989
Db 1734 GGGACCCAGTGGCCCGGATCTGCTGCTGAATGGCAGCCCATCCAGTAGCTCGC 1793
QY 990 GluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThr 1009
Db 1794 TCCACCTCCGAGGCC-----GGCGTGCCTGAGCTTCCACATCCAGGATGCC 1838
QY 1010 ThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSer 1029
Db 1839 CTGCCGAGGACCATGACCTACACCTGCTAGCTGAGATGCTTGGGCGAGGTGCC 1898
QY 1030 CysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGly 1049
```

```
Db 1899 TGCAGCGCCTGGGTACCGTCCAT----- 1922
QY 1050 GlnSerHisArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArg 1069
Db 1923 -----GAAAGAAAGAGTAGCAGGAGAGTAGTACCTTCTGCTGTGGCTCCAGCAAG 1976
QY 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGlyArgLeu 1089
Db 1977 CCCACTGCACCCATCTTCTGCGAGGCTCTCTGATCTCAAGATCATGGATGGAAGCCAG 2036
QY 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsn 1109
Db 2037 GTCACTATGACTGTCAAGTGTGAGGAATCCACCCCTCAAGTCACTGTGCTGCAAT 2096
QY 1110 GlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
Db 2097 GGAATGAGATCCAAGAGTCAAGAGACTTCCACTTTGAACAGAGAGAACT---CAGCAC 2153
QY 1130 SerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThr 1149
Db 2154 AGCCTTGGATCCAGGAAGTGTCCCGAGGACACAGGCGACGTACACTCGGAGCCTGG 2213
QY 1150 AsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
Db 2214 AACAGCGCTGGAGAGTCCGACCCAGGCGCTGCTACGGTACAAGAGCCCTCAGCATGGC 2273
QY 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
Db 2274 ACCAGCCCTGGTTCATCAGTAAGCCTGCTGCTAGTACAGCCCTCCCTGGGCCAGAGTGC 2333
QY 1190 ArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAspAsn 1209
Db 2334 CTATCTCTCTGGCCATAGCTGTGACCCCTTCTACCGTGCACCTGGCTCAGAGATGGC 2393
QY 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAla 1229
Db 2394 AAAGCCCTCTGCAAAAGACACTGGCCACTTCGAGGTGCTTCAGAAATGAGGAGCTTTCACC 2453
QY 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLys 1249
Db 2454 ---CTGTTCTTAAAGAGGTGACGCCCTGGCATGCCGCCAGTATGAGATCCTGCTCAAG 2510
QY 1250 AsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis 1269
Db 2511 AACCGGTTGGCGAATCAGTTCGCGAGTGTCTACTGATGTCTACAGAACAGCTCTGCCAGA 2570
QY 1270 GlnIlePro-----ProMetSerValArgPro----- 1279
Db 2571 GCCCTTCCAGGGGGAGGAGGAGCCTGCCAGCTGCCAGGACCTCTGTGGTGGAGGAGTTGGT 2630
QY 1280 -----SerGlySerArgTyrGlySerLeu 1287
Db 2631 GCTGATGGTGGTGTAGTGACCGCTATGGGTCCCTG 2666

RESULT 5
US-09-954-456-522
; Sequence 522 Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
```


Percent Similarity:	33.71%	Conservative:	124
Best Local Similarity:	22.96%	Mismatches:	435
Query Match:	8.09%	Indels:	330
DB:	10	Gaps:	42

US-09-818-990B-2 (1-1320) x US-09-880-107-3371 (1-5926)

Qy	195	AsnSerSerPheSerAspLeuSerGluArgGluArgSerValProPhePro	214
Db	18	AGTTGACGACAACTTCAGGAGCTTCCCGCGCAGAGACTTCAGCAGCCGCTTCTGCTGCCA	77
Qy	215	-----lleProAlaAspThrArgAspAsnGluValAsnHisAlaLeuGluGln	231
Db	78	CTGCCCCAGTTGCCCAACAAACACAGAGAACCGTGCACCATGGG	125
Qy	232	GluAlaLysArgGluAlaGluGlnAlaAlaSerGluAlaGlyGlyAspThr	251
Db	126	-----GATGTGAAGCTGGTTGCCCTCGTCGCACACATT--TCCAAAACCTCC	167
Qy	252	ProGlySerSerProSerSerLeuTyrTyrGluLupProLeuGlyGlnProProArgPhe	271
Db	168	CTCAGTGTGATCCCTCAGAGTT---GACTCCATGCCCTGCAGAGCCCTGCTTTC	224
Qy	272	ThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGlnLeuAspCys	291
Db	225	ATTTTGGCCCTCGGAACCTCTGCATCAAGAGGAGCCACCGCCAAGTTCGAAGGGCGG	284
Qy	292	ValValGlyIleProProGlnValArgTrpTyrCysGluGlyLysGluLeuGluAsn	311
Db	285	GTCCGGGGTTTACCAGAGCCCGAGGTGCATGTCACAGAAAGCGGCAACCCATC---ACC	341
Qy	312	SerProAspIleHisIleValGlnAlaGly-----AsnLeuHisSerLeuThrIleAla	329
Db	342	AGCGGGGGCCCTTCCTGCTGGATTGCGGCATCCGGGGGACTTTCAGCCCTTGATCAT	401
Qy	330	GluAlaPheGluGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThr	349
Db	402	GCTGTCCATGAGGAGGACAGGGGAAAGTATACCTGTGAAGCCACCAATGGCAGTGGTGT	461
Qy	350	AspSerThrSerAlaGluIleTyrIleGluGlyValSerSerSerAspSerGluGlyAsp	369
Db	462	CGCCAGGTGACGTGGAGTTGACAGTAGAAGGA--AGTTTTCGGAAGCAGCTTGGTCAG	518
Qy	370	Pro-----AsnLysGluGluMetAsnArgIleGlnLysProAsnGlnValSerSerPro	387
Db	519	CCTGTTGTTTCCAAAACCTTAGGGGATAGATT--	551
Qy	388	ProThrThrSerAlaValIleProProAlaValProGlnAlaGlnHisLeuValAlaGln	407
Db	552	---TCAGCTTCACGACGTG-----GAG	569
Qy	408	ProArgValAlaThrIleGlnGlnCysGlnSerProThrAsnTyrLeuGlnGlyLeuAsp	427
Db	570	ACCGTCTAGCATCTGGGGGAGTGC-----	596
Qy	428	GlyLysProIleAlaAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSer	447
Db	597	-----CCACCAAGATTGTGTACCAAGCTGGCGCGAGTTGTTGTCACAA	638
Qy	448	GluGlyGlnLeuValValPheGluCysArgValLysGlyAlaProSerProLysValGlu	467
Db	639	GAAGACAGATGGGACCATCTCTCGCAGATCATCTGGCGGGCCCCAACCGCAGGTCCAC	698
Qy	468	TrpTyrArgGluGlyThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLys	487
Db	699	TGGCTCAAGGGAATGTTCCACTGTCAG-----CCGAGTGCCTGTG-----	740
Qy	488	ProArgSerMetAlaGluProGluGluIleCysThrLeuValIleAlaGluValPheAla	507
Db	741	-----TCTGTGTCGAGAAGAACGCATGCAGGTCTCTGGAATCCTGGATCAACCAA	794
Qy	508	GluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIle	527

Db	795	GATGACGTGGGAGTGTACACGTGCCCTGGTGGTGAACGGGTGGGGAAGGCCCTCGATGTC	854
Qy	528	AlaGlnLeuHisValArgGly	534
Db	855	GCTGAACCTTTCCATCCACGAGTTGGACAGCTGGCAATAGGTCATTTGTCAGAGAACACAAA	914
Qy	535	-----AsnGluAspLeu	541
Db	915	GCCACCAATTTCAGATGTCAGGAAGAGGTGACCAATGTAACTCAAGAGGAGTCGAAGCTG	974
Qy	542	GlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaLalleGluProGlnProSer	561
Db	975	GACATCTGGAGGCTGCAGCAAAACAGAACTGCTCCAGCCCCAGAGAGGTGGCTTC	1034
Qy	562	ProPro-----HisSerGluProProSerValGluGlnProProLys---ProLys	577
Db	1035	CCACCTTGGGCTGCAACACCCAGCCT-----CAGCCCCCAAGGAGGTCCAG	1082
Qy	578	LeuGluGlyValLeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArg	597
Db	1083	CTGGAGTCATGC-----AAGGACTCGCCAGACGGCC-----	1115
Qy	598	ValHisPheAsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyVal	617
Db	1116	-----CCGCAGACCCCGGTCCTTCAGAAAGACTTCCAGCTCC-----	1151
Qy	618	ValThrThrArgGlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThr	637
Db	1152	ATCACCTCGAGCGCGCAGAGTTCAG-----	1178
Qy	638	LysThrProGluProSerPhePro-----ValLysGluProProProValLeuAla	654
Db	1179	-----CCGGACCAACAGACACCAGCCCTGGGGGTCTCTACCT-----	1217
Qy	655	LysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGln	674
Db	1218	-----TCTGGA	1223
Qy	675	HisGlnLeuGlnAsnProProProSerSerProLysGluPhePheProPhe**MetThrVal	694
Db	1224	GAAGAGAGGAAGAGGCCAGCTCTCCCTCCAGCCACTTCCCCCAGCAGCAGCCTGGC	1283
Qy	695	LeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysAlaProSer	714
Db	1284	CTGGGGAGCCAA-----GATGTTGTGACGAGGTGCTAACAGGAGAATCCCCATG	1334
Qy	715	SerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnThrAla	734
Db	1335	GAGGGCCAGAGGATTTCAGCATTTCCCAAAATTT-----	1367
Qy	735	AlaThrValAlaProSerSerProValPheThrLeuSerSerThrProGlnThrIle	754
Db	1368	-----GAGAGCAAGCCCCCA-----	1382
Qy	755	GlnArgThrValSerLysGlu-SerLeuLeuValSerHisProSerValGlnThrLysSe	774
Db	1383	-----AGCCAGGAGTCAAGGAATCAAACTGTCAGTTCAGATGTCAGATT	1430
Qy	774	rProGlyGlyLeuSerIleGln-AsnGluProLeuProProGly-----ProThrGluP	792
Db	1431	TCCGGGATTCCAAAGCCTGAAGTGGC-----CTGGTTCTTGGGAAGGACACCCCGT	1484
Qy	792	roThrProProProPheThrPheSerIleProSerGlyAsnGlnPhe-----GlnProA	810
Db	1485	AGACA-----GGAAGGCAGCATTTGAGTTTATGAA	1514
Qy	810	rgCysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheL	830
Db	1515	GATGCTGGCTCCCATTTACCT-----	1534
Qy	830	euSerSerValLeuProSerLeuProAlaIleProPro-ThrAsnAlaMet***LeuPro	849
Db	1535	-----CTGCCTGCTGAAGAGCCCGGAC-----	1556

Qy	1210	GluthrIleProCysThrArgGluAlaGlySerMetHisGlnAspThrThrGlyTyrAla	1229
Db	2394	AAAGCCCTTCGCAAGACACTGGCCACTTCAGGTGCTTCAAGAATAGGACGCTGTTCAACC	2453
Qy	1230	CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLys	1249
Db	2454	--CTGGTTCATANAAGGTGCAGCCCTGGCATCGCGGCAAGTAGTGAGATCCTTGCTCAAG	2510
Qy	1250	AsnGluAlaGlyIleValSerCysThrAlaArGLeuAspIleTyrAlaGlnTrpHisHis	1269
Db	2511	AACCGGTTGCCAATGCAGTTGCCAGGTCTCACTGATCTACAGAACAACCTCTGCCAGA	2570
Qy	1270	GlnIlePro-----ProMetSerValArgPro-----	1279
Db	2571	GCCTTTCACGGGGAGGAGAGCTGCCAGCTCGAGGAGCTCTGTGTGGAGGAGTTGGT	2630
Qy	1280	-----SerGlySerArgTyrGlySerLeu	1287
Db	2631	GCTGATGTGGTGGTAGTGACCGCTATGGGTCCCTG	2666
RESULT 7			
US-09-880-192-32			
; Sequence 32, Application US/09880192			
; Patent No. US20020077470A1			
; GENERAL INFORMATION:			
; APPLICANT: Walker, Michael G.			
; APPLICANT: Volkmueth, Wayne			
; APPLICANT: Klingler, Tod M.			
; APPLICANT: Azimzai, Yalda			
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION			
; FILE REFERENCE: PB-0009-1 CIP			
; CURRENT APPLICATION NUMBER: US/09/880,192			
; CURRENT FILING DATE: 2001-06-12			
; NUMBER OF SEQ ID NOS: 62			
; SOFTWARE: PERL Program			
; SEQ ID NO 32			
; LENGTH: 1833			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; OTHER INFORMATION: Incyte ID No. US20020077470A1 3795510CT1			
US-09-880-192-32			
Alignment Scores:			
Pred. No.:	5,75e-27	Length:	1833
Score:	523.50	Matches:	100
Percent Similarity:	66.08%	Conservative:	50
Best Local Similarity:	44.05%	Mismatches:	74
Query Match:	7.60%	Indels:	3
DB:	10	Gaps:	2
US-09-818-990B-2 (1-1320) x US-09-880-192-32 (1-1833)			
Qy	1096	ValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuPro	1115
Db	673	GTGAGTGACTGCCAGCTCCTGATGTGTATGATATAATGAAGAACAGTTCAATCA	732
Qy	1116	AspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspPro	1135
Db	733	GATGATTTCACAAAATGATAGTGTCTGAGAGGGTCTTCATCTCACTCATCTTTGAGTA	792
Qy	1136	LeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsn	1155
Db	793	GTCAGAGCTTCAGATGCAGGGCTTATGCATGTGTGCCAAGAAATAGACGAGAGACCC	852
Qy	1156	SerpHeSerLeuGluLeuSerValValAlaLaLysGluValLysAlaProValIleLeu	1175
Db	853	ACCTTTCAGCTGTGCAGCTGGATGTCTTCGAAAAAGACATAAAAGACACCAATGTTTTATC	912
Qy	1176	GluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgVal	1195
Db	913	TACAAACACACAGAGCAAAAGAGTTTATAGAGGAGATTCAGTGAACCTAGAAATGCCAGATC	972

```
Qy 1196 IleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThr 1215
Db 973 TCGGCTATACCTCCACCAAGCTTTCTCGAAAGAATAATGAATGGTACATTCAAC 1032
Qy 1216 ArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnPro 1235
Db 1033 ACTGACCAATAAGCTTATATCAAGATAACACTGGAAGAGTTACTTTACTGATAAAGAT 1092
Qy 1236 AlalysSerAspAlaGlyTyrThrLeuSerAlaLysAsnGluAlaGlyIleVal 1255
Db 1093 GTAAACAAGAAGATGCTGGGTGTATCTGTGCAGCAGTAAATGAAGCTGGAGTGACT 1152
Qy 1256 SerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIleProProPro--- 1274
Db 1153 ACATGTACACAAAGATTAGAGTTACGGCAGTCCAAACCAACTCTTCCAGCTCCTTAAG 1212
Qy 1275 ---MetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAsp 1293
Db 1213 CAGTTACGGGTTCGACCAACACTTCAGCAAAATATTAGCACTTAATGGGAAAGTTTGAAT 1272
Qy 1294 IlePheSerAlaPheSerSerMetGluSerThrMetValTyrSerCysSerSerArgSer 1313
Db 1273 GTAAACAAGCTTTTAACCCA---GAAGAGAAATTTTCAGCGTTTGGCAGCTCAATCTGGA 1329
Qy 1314 ValValGluSerAspGluLeu 1320
Db 1330 CTCATGAAAGTGAAGAACTT 1350
RESULT 8
US-10-077-130-6
; Sequence 6, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6
Alignment Scores:
Pred. No.: 1.4e-24 Length: 23907
Score: 508.50 Matches: 302
Percent Similarity: 35.03% Conservative: 181
Best Local Similarity: 21.90% Mismatches: 487
Query Match: 7.38% Indels: 412
DB: 9 Gaps: 52
US-09-818-990b-2 (1-1320) x US-10-077-130-6 (1-23907)
Qy 122 ProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyrCys 141
Db 14984 CCGAGAGCTCTCTCGGGGTGAGCTGGACGATGCTTCCGCGGGGCTCCCGTCGGGTGC 15043
Qy 142 SerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluLeu 161
Db 15044 ACCGGCTCTTCGCGACCAAAAGTCCGGCTGAAG----- 15076
Qy 162 SerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 15077 -----TTTCAGATGAGGAGCTCTTCCTGTAGTGCAGAGGGGCCCTGCAGAGCCAG 15127
```

```
Qy 182 LysSerLysLeuGluSerGlnAsn-----LysValMetGlnGluAsnSerSerSerPhe 199
Db 15128 AGGAGCCCGCGAGCTGGCAGACATACCAGGAAGATGAGCATTTTCATTCGTCACCT-TTT 15186
Qy 200 SerAspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThr 219
Db 15187 GAGCGGCTCACTAGAGCCGCCAG-----ValAsnHisAlaLeuGluGlnGlnGluAla 233
Qy 220 ArgAspAsnGlu-----ValAsnHisAlaLeuGluGlnGlnGluAla 233
Db 15220 CGCTTCCAGGAGATGTTTCCACACTGGGCATTTGGGTGGAGATCAAGCTGGTGGACAG 15279
Qy 234 LysArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGly 253
Db 15280 GGGCTCGGAGGGTAGAGATGTCATCAGCAA----- 15312
Qy 254 SerSerProSerSerSerLeuTyrTyrGluGluProLeu-----GlyGln 267
Db 15313 GAGACTCTCGCCCTGGTGGCTCCAGAGCAATTGCCAGCCTACTACTTCTGACGCT 15372
Qy 268 ProProArgPheThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGln 287
Db 15373 GCCCAGTGTTCCTGACTGAGTTGCAGAACCAAGATGTCAGATGGGTATCTCTGTGAGC 15432
Qy 288 LeuAspCysIleValValGlyIleProProGlnValArgTyrTyrCysGluGlyLys 307
Db 15433 TTTGACTCGGTGTGACAGGTCAGCCCATGCCAGTGTGCGCTGTTCAGAGATGGGAAG 15492
Qy 308 GluLeuGluAsnSerProAspIleHisIleVal-----GlnAlaGlyAsnLeu 323
Db 15493 TTGTTGGAGGAG-----GATGATCACTATCATGATTATGAAGACCAACAGGGTGGC 15543
Qy 324 HisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCysPheAla 343
Db 15544 CATCAGCTCATCATCACAGCCGTGGTCCACAGACAGATGGGGCTGTACCGCTGCTGGCC 15603
Qy 344 SerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGlyValSerSer 363
Db 15604 GAGAACAGCATGGGTGCTCTCTCCACCAAGCTCAGCTCCGTGTGGAC---TTGACAAGC 15660
Qy 364 SerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLysProAsnGlu 383
Db 15661 ACAGACTATGAC----- 15672
Qy 384 ValSerSerProProThrThrSerAlaValIleProProAlaValProGlnAlaGlnHis 403
Db 15673 -----ACTGCAGCA-----GATGCCAGGAGTCTCATCTCTAC 15705
Qy 404 LeuValAlaGlnProArgValAlaThrIleGlnGln-----CysGlnSerProThrAsn 421
Db 15706 TTCAGTGCCTCAAGGCTACCTGTCCAGCGGGAGGAGGAGGAGGAGTCCACCACTGAT 15765
Qy 422 TyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaProValPheThrLysMetLeu 441
Db 15766 -----GAGGGCCAG-----CTGCCCCAGGTTGGTGGAGGAGCTG 15798
Qy 442 GlnAsnLeuSerAlaSerGluGly---GlnLeuValValPheGluCysArgValIleGly 460
Db 15799 AGAGACCTCCAGGTGGGCCCTGGCCACAGCTGCCAAGTTCCAGCTCAAGGAGGAGG 15858
Qy 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
Db 15859 TACCCTGCTCCAGATATCTGTCTCAAGATGCGCAGCCCTGACCGCATCTGCCAC 15918
Qy 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu 500
Db 15919 ATCCGCATGACTGCGCAAGAAG-----ATCCTGCACACCTGT 15954
Qy 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
Db 15955 GAGATCATCTCCGTCACCGGGAGGACTCTGGCCAGTATGACGCTATATCAGCAATGCC 16014
Qy 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
```

Db	16861	ATCAGCGTCGGGGAAGGCCCATATGTGTGAGGTCTCTGGATGACAGCCACCCACCTCGCGTGG	16929
Qy	847	*LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSe	867
Db	16921	CTTG-----TCGCACCAAGCCCAAGTCCAGCCCC--TCAGCGCAGGGCTGGGTGTC	16973
Qy	867	rPro-----	868
Db	16974	ACCAGCCTTACCTGGACAGGAGGCTCAAGCTGTCACTGTAGTGGGGGCGCTGAGGCCCC	17033
Qy	869	-----GlnProValAsnAspAspAsnIleArgGluThrLysAsnAlaValI	884
Db	17034	TGAGTTCCCTGGGAGGCTGTGTCTGAAGACCAATACAAGCAAGGCTGAGCTCTGTGAT	17093
Qy	884	eArgAspLeu-----	887
Db	17094	CCAGGAGCTGTGAGTTCTGTGACGAGGCCCTTCGTGGAGGAGCTGCAGTTCTTCAGAGCCA	17153
Qy	888	-----GlyLysLys	890
Db	17154	CCACCTGCAGCACCTGGAGCGCTGCCCCACAGCTGCCCATAGCCGTGGCGCGCCAGAAGGC	17213
Qy	890	silThrPheSerAspValArgProAsnGlnGlnGluThrLysLysSerPheGluG	910
Db	17214	AGTCATCTTCGCGAATGTGCGG-----GACATGCGCGCTCTCCACAG	17255
Qy	910	nArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspAs	930
Db	17256	CAGCTTCTCGAGGAG-----TTGCAGCAGTGCACACGGGACGAGCT--GG	17302
Qy	930	pGluIleGlnHisaspGlu-----IleProThrGlyLys	941
Db	17303	CCATGTGCTTCATCAAGAACCAGCGCGCTTTTGACGAGTACTGTGGATTCTCTGGTGGGCG	17362
Qy	941	sCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGlu-----	958
Db	17363	GTGTGCAGGCTGAGTCGTGTGTGTCGACACGGCCATCCAGAGGTCTACAGAAATACG	17422
Qy	959	-----GlySerProValThrPheThrCysLysI	968
Db	17423	CGGAGGAGCCCTGTTGCGAGGGGACCCTCTCAGCCCCCGCCACCCTCTGCAGCACT	17482
Qy	968	eValGlyIle-ProValProLysValTyrTrpPhe-----LysAspGlyLysG	984
Db	17483	ACCTGGACGACGCTGGAGCGGTGTCAGCGCTACACGCGCTTGCTGAAGGAGTTGATCC	17542
Qy	984	nIleSerLysArgAsn---GluHisCysLysMetArgArgGluGlyAspGlyThrCysS	1003
Db	17543	GCAACAGGCGCGGAACAGACAGAACTCGCGCTCTCGAGCGCCATGCCGTGGTGT	17602
Qy	1003	er-----LeuHisIleGluSerThrThrSerAspAsp-	1013
Db	17603	CTGCCCTGCCACAGCGCGCTGAGAACAGCTGCAGCTGTCCCTCATGAGAACTACCCAG	17662
Qy	1014	-----AspGlyAsnThrThrIleMetAlaAla	1023
Db	17663	GCACCCCTGGAGGCCCTGGGCGAGCCATCCGCGACGGCCACTCATCGTGTGGGAGGTG	17722
Qy	1023	snProGlnGlyArgIleSerCysSerGlyHis-----L	1034
Db	17723	CACCGGGGGCGCGATGCTCCCTGGAAGGCCAACACCGTCACGTGTCTCTTCGCCAACCC	17782
Qy	1034	eMetValGlnSerLeuProfileArg---SerArgLeuThrSerAlaGlyGlnSerHis	1053
Db	17783	ACCTGGTAACTGCAAGCCCGCGGAGACTCCCGCACCGATACCGTCAGCTACGTGTCC	17842
Qy	1053	rgGlyArgSerArgVal-----	1058
Db	17843	GGAACATGATGAAGCTGAGCACCATCGACTGAACACCAAGGTGGAGGGGATGACCGCG	17902
Qy	1059	-----GlnGluArgAsp-----LysGluProLeuGlnGluArg-	1069
Db	17903	CTTCGAGGTGTGGCAGGAGCGGGAGGACTCGGTGCGGAAGTACCTGTGTCAGGACCGGA	17962


```
Qy 1069 ----- 1069
Db 17963 CAGCCATTATCAAGAGCTCGTGGGTGAAGAGATCTGTGGCATCCAGCAGCGTCTGGCCC 18022
Qy 1070 -----pHeAArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluG 1087
Db 18023 TGCGTGTGTGGCGGCCCGGACTTTGAAGAGGAGCTGGCCGACTGCACAGCCGAGCTGG 18082
Qy 1087 lYArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpL 1107
Db 18083 GTGACAGCTCAAGCTGCCTGCCGTGACGGGACACACCCAGCCTGTATCAGCTGGT 18142
Qy 1107 euLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrG 1127
Db 18143 ACAAGATGGGAACAGCTGCAGGTGGACCCACCACATCCTATTGAAGACCCCTGATG 18202
Qy 1127 lYValHisSerLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysI 1147
Db 18203 GCTCGTGTGCACATCATCTGTGACACGCTGACCGGTGTGGACTCTGGCCAGTACATGTGCT 18262
Qy 1147 leAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLys- 1166
Db 18263 TCGCGGCCGAGCCCGCTGGC-----AACTGCAGTACCCTGGGCAAGA 18304
Qy 1167 --GluValLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluG 1186
Db 18305 TCCTGGTGCAAGTCCACCACCGGTTCTGTGAACAGGTCCGGGCCCTCACCCCTTTGTGGAGG 18364
Qy 1186 lYHisProValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpL 1206
Db 18365 GAGAGGAGCGCCAGTTCACCTGCACCATCGAAGGCGGCCCTACCCGCGAGATCAGGTGGT 18424
Qy 1206 ysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrT 1226
Db 18425 ACAAGAGCGGGCCCTGCTGACCACTGGCAACAAGTTCAGACACTAGTGGAGCCTCGCA 18484
Qy 1226 hrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrL 1246
Db 18485 GCGCGCTGCTAGTGTGTGTATCCGGCGGCGCCAGCAAGGAGGACCTGGGCTCTACGAGT 18544
Qy 1246 euSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIle 1263
Db 18545 GTGAGCTGGTGAACCGCTGGGCTCCGCGGGGCTAGTGGGAGCTGCGCAT 18597
RESULT 9
US-10-077-130-4
; Sequence 4, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 59079 and 12599, protein Kinase Family
; FILE REFERENCE: MF12001-047(PICP(M))
; CURRENT APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4
```

```
Alignment Scores:
Pred. No.: 1,41e-24 Length: 24120
Score: 508.50 Matches: 302
Percent Similarity: 35.03% Conservative: 181
Best Local Similarity: 21.90% Mismatches: 487
Query Match: 7.38% Indels: 412
DB: 9 Gaps: 52

US-09-818-990B-2 (1-1320) x US-10-077-130-4 (1-24120)
Qy 122 ProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyrCys 141
Db 15055 CCGAGAGCTCTCTGGGGTGGAGTGGAGTGCCTCCGCGCGGCTGCCGCTGGCTGC 15114
Qy 142 SerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGluLeu 161
Db 15115 ACCGGCTTCTCCGACCAAAAGTCCGGCTGAAG----- 15147
Qy 162 SerSerLysPheLysSerHisSerSerLysArgIleArgProAlaGlaCysLysAsnHis 181
Db 15148 -----TTTCAGATCAGGAGCTCTCTCAGTGCAGAGCGGCTCCAGAGCCAG 15198
Qy 182 LysSerLysLeuGluSerGlnAsn-----LysValMetGlnGluAsnSerSerPhe 199
Db 15199 AGGAGCGCGGAGTGGCAGACATACCGGAAGATGAGCATTTTCATCTGCATCCGT-TTT 15257
Qy 200 SerAspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThr 219
Db 15258 GAGCGCTCACTGAGGCGCCGCGCAG-----GCAGTAACT 15290
Qy 220 ArgAspAsnGlu-----ValAsnHisAlaLeuGluGlnGlnGluAla 233
Db 15291 CGCTTCAGGAGATGTTGCCACACTGGGCATTTGGGTGGAGATCAAGCTGGTGGACAG 15350
Qy 234 LysArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGly 253
Db 15351 GGGCTCGGAGGTAGAGATGTGCATCAGCAA----- 15383
Qy 254 SerSerProSerSerLeuTyrTyrGluGluProLeu-----GlyGln 267
Db 15384 GAGACTCTCCCTGCTGGTGGCTCCAGAGCCATTGCCAGCCACTGCCACTTCTGAGCGCT 15443
Qy 268 ProProArgPheThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGln 287
Db 15444 GCCCAGTGTTCCTGACTGAGTTGCAGAACCAAGATGCGATGGTATCCTCTGTGAGC 15503
Qy 288 LeuAspCysIleValValGlyIleProProGlnValArgTrpTyrCysGluGlyLys 307
Db 15504 TTTGACTCGGTGTGACAGGTGACGCCATGCCAGTGTGCGCTGGTTCAGAGATGGGAAG 15563
Qy 308 GluLeuGluAsnSerProAspIleHisIleVal-----GlnAlaGlyAsnLeu 323
Db 15564 TTGTTGGAGGAG-----GATGATCACTACATGATTATGAAGACCAACAGGGTGGC--- 15614
Qy 324 HisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCysPheAla 343
Db 15615 CATCAGCTCATCATCACAGCGGTGGTGGCCAGACACATGGCGCTACCGCTGCTGGCC 15674
Qy 344 SerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGlyValSerSer 363
Db 15675 GAGAACAGCATGGGTGCTCTCCACCAAGGCTGAGCTCGCTGTGGAC---TTGACAAGC 15731
Qy 364 SerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLysProAsnGlu 383
Db 15732 ACAGCATATGAC----- 15743
Qy 384 ValSerSerProThrThrSerAlaValIleProProAlaValProGlnAlaGlnHis 403
Db 15744 -----ACTGCAGCA-----GATGCCACGGAGTCTCTCATCTCTAC 15776
Qy 404 LeuValAlaGlnProArgValAlaThrIleGlnGln-----CysGlnSerProThrAsn 421
Db 15776 -----GATGCCACGGAGTCTCTCATCTCTAC 15776
```


Db 15777 TTCAGTGCACCAAGGCTACCTGTCCACCGGGAGCAGGAGGAACAGAGTCCACCACTGAT 15836
QY 422 TyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaProValPheThrLysMetLeu 441
Db 15837 -----GAGGGCCAG-----CTGCCCCAGGTGGAGAGCTG 15869
QY 442 GlnAsnLeuSerAlaSerGluGly----GlnLeuValValPheGluCysArgValLysGly 460
Db 15870 AGAGACCTCCAGTGGCCCTGGCACACAGCTGGCCAAAGTTCACGCTCAAGGTGAAGGC 15929
QY 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuLeuGluAspSerProAsp 480
Db 15930 TACCTGTCCTCCAGATATACCTGGTCAAGATGGCAGCCCTCAGCCGATCGCCAC 15989
QY 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu 500
Db 15990 ATCCGGATGACTGGCAAGAG-----ATCTGCGCACCCCTG 16025
QY 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
Db 16026 GAGATCATCTCCCTACCCCGGAGGACTCTGGCCAGTATGCAGCCTATATACGAATGCC 16085
QY 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
Db 16086 ATGGGTGCTGCCTACTCGTCTGCCCGGCTGCTGGTTGAGGC----- 16127
QY 541 AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnPro 560
Db 16128 -----CCT 16130
QY 561 SerProHisSerGluProProSer---ValIleGlnProProLysProLysLeuGlu 579
Db 16131 GATGAGCCAGAGAAGCCCTGCATCAGATGTGCATGAGCAGCTGTGTCG 16181
QY 580 GlyValLeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHis 599
Db 16182 -----CCCCAATGCTGGAGAG----- 16199
QY 600 PheAsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThr 619
Db 16200 TTCACCCCAAGAAGTGAAGAAGCTCCAGCATCACCTCTCTGTGAAGTA----- 16253
QY 620 ThrArgGlnThrArgProAspSer***-----GlnGluArgPheAsnGly 634
Db 16254 -----GAAGGACGCGCGGCCACCGTGCATGCTGCTCAGGAGGAGCTGAGAGGC 16307
QY 635 GlnAlaThrLysThrProGlu---ProSerPheProValLysGluProProValLeu 653
Db 16308 GTCTGTGGATTGGCCCTTGACACACCGGGCTACACCGTG----- 16346
QY 654 AlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGlu 673
Db 16347 -----GCCAGCTTGGCCACAG---CACAGCCTGGTCTCTGCTGGAC 16385
QY 674 -----GlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe 690
Db 16386 GTGGCGCGGACGACCCAG-----GGC 16406
QY 691 ***MetThrValLeuAsnSerAsnAlaProProAlaValThrSer***LysGlnVal 710
Db 16407 ACCTACACATGCTATGTCACACACGCTGCGCGGCAGGCCCTCTGCTCCGCGCCTGCAC 16466
QY 711 LysAlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThr 730
Db 16467 GTCTCGGGCCCTGCTAAGGTGGAGGACGAGGAAGTGAAGAGCGCTGATTTCACCT 16526
QY 731 AsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer----- 748
Db 16527 TTCCTGCGGGGACCACACAGCCATCTCAGACAGGGGTGGAAACTGCGAGTTTGCT 16586
QY 748 ----- 748
Db 16587 GACCTTGGTGGGACAGAGAAAGAGCCCTCTGGCTGCCAAGGAGGCCCTCGGCCACCTG 16646

QY 749 -----SerThrProGlnThrIleGlnArgThrValSerLysGluSerLeu 763
Db 16647 TCCTCGCTGAGTGGCACAGAGGAGTTCTTCGAGAACTGACCTCCAGATCACTGAG 16706
QY 764 LeuValSerHisProSerValGlnThrLys-----SerProGlyGlyLeuSerIleGln 781
Db 16707 ATGTATCGGCCCAAGATCAGCGAGGCAAGCTGCAGGTGCGCCGAGGTGACAGTATGAG 16766
QY 782 AsnGluProLeuPro-----ProGlyProThrGluProThrPro-ProProPheTh 798
Db 16767 GACTCCAAGACACACTCTGCATCCCCCGCCATGGCGGATCAGCGCCATCTCCAGCATC 16826
QY 798 rPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSe 818
Db 16827 CAGGAGTCTTCCTCAGAGTCAG-----AGGACGGCATGCCCGGCGGAGATC 16874
QY 818 rProThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuPr 838
Db 16875 TTTGACATCT---ACGTGGTCACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16931
QY 838 oAla-----IleProProThrAsnAlaMet** 847
Db 16932 ATCAGCTGCGGAAAGGCCAGTATGTGGAGTCTCTGATGATGAGCCACCATCGCTGG 16991
QY 847 *LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSe 867
Db 16992 CTGTG-----TCGCGACCAAGCCACCAAGTCCAGCCCC---TCAGGCGAGGCTGGTGTCT 17044
QY 867 rPro----- 868
Db 17045 ACCAGCTACCTGGACAGGAGCTCAAGCTGTCACTGAGTGGGGGCGCTGAGGCC 17104
QY 869 -----GlnProValAsnAspAspAsnIleArgGluThrLysAsnAlaValI 884
Db 17105 TGAGTTCCCTGGGAGGCTGTGTGAAGACGAATACAAGCAAGGCTGAGCTGTGTAT 17164
QY 884 eArgAspLeu----- 887
Db 17165 CCAGGAGCTCTGAGTCTCAGCAGGCGCTTGTGGAGAGCTGCATCTCTGACAGACCA 17224
QY 888 -----GlyLysLys 890
Db 17225 CCACCTGCAGCACCTGGAGCGCTGCCCCACGTGCCATAGCCGTGCCGCGCAGAGGC 17284
QY 890 sIleThrPheSerAspValArgProAsnGlnGlnGluThrLysLysSerSerPheGlu 910
Db 17285 AGTCATCTTCGCAATGTGCGG-----GACATCGCGCGCTTCCACAG 17326
QY 910 nArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspAs 930
Db 17327 CAGCTTCCTGCAGGAG-----TTGACAGCTGCGACGACGACGACGCT-GG 17373
QY 930 pGluIleGlnHisAspGlu-----IleProThrGlyLys 941
Db 17374 CCATGCTCTTCATACAAGAACCGCGCGCTTTGAGCAGTACCTGGAGTTCTCTGTGGG 17433
QY 941 sCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGlu----- 958
Db 17434 GTGTGAGGCTGAGTGTGCTGCTGACAGCGCCATCCAGGAGTTCTACAGAAATACG 17493
QY 959 -----GlySerProValThrPheThrCysLysI 968
Db 17494 CGGAGGAGGCCCTGTTCGACGGGACCCCTCTCAGCCCCCCCCACCATCTCTGCAGCACT 17553
QY 968 eValGlyIle-ProValProLysValTyrTrpPhe-----LysAspGlyLysG 984
Db 17554 ACCTGGAGCAGCCAGTGGAGGGGTGCAGCGCTACCAGGCCCTTCTGTAAGGAGTTGATCC 17613
QY 984 lnIleSerLysArgAsn---GluHisCysLysMetArgArgGluGluArgGlyThrCys 1003
Db 17614 GCAACAAGGGCGGGAACAGACAGAACTGCGGCTGCTGGAGCAGGCGCTATCCCGTGTGT 17673

```
Qy 1003 er-----LeuHisIleGluSerThrThrSerAsp- 1013
||
|||
Db 17674 CTGCCCTGCCACAGCGCTGAGAACAGCTGCAGCTCCTCATGGAGAACTACCCAG 17733
|||
Qy 1014 -----AspGlyAsnTyrThrIleMetAlaAla 1023
|||
Db 17734 GCACCTGGAGGCCCTGGGCGAGCCCAATCCGCCAGGGCCACTTCATCGTGTGGAGGGTG 17793
|||
Qy 1023 snProGlnGlyArgIleSerCysSerGlyHis-----L 1034
|||
Db 17794 CACCGGGGCCCGCATGCTCCCTGGAGGCCACACCCGTCACGTGTTCTCTTCGCGCAACC 17853
|||
Qy 1034 euMetValGlnSerLeuProIleArg-----SerArgLeuThrSerAlaGlyGlnSerHisA 1053
|||
Db 17854 ACCTGGTAATCTGAAGCCCGCGGAGAGACTCCCGCACCGATACCGTCAGCTACGTGTTC 17913
|||
Qy 1053 rgGlyArgSerArgVal----- 1058
|||
Db 17914 GGAACATGATGAAGCTGAGCAGCATCGACCTGAACGACCGAGGTGGAGGGGATGACCGCG 17973
|||
Qy 1059 -----GlnGluArgAsp-----LysGluProLeuGlnGluArg- 1069
|||
Db 17974 CCTTCGAGGTGTGCAGGAGCGGGAGGACTCGGTGCCCAAGTACTGCTGCAGGCACGGA 18033
|||
Qy 1069 ----- 1069
|||
Db 18034 CAGCCATTATCAAGAGCTCGTGGTGAAGGAGATCTGTGGCATCCAGCAGCGTCTGGCCC 18093
|||
Qy 1070 -----PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluG 1087
|||
Db 18094 TGCCTGTGTGGCGCCCGCGACTTTGAAGAGGAGCTGGCGCATGCACAGCGAGCTGG 18153
|||
Qy 1087 lyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpL 1107
|||
Db 18154 GTGACAGCTCAAGCTGCCTCGCGCTGAGCGGCACACCCAAAGCTCTCATCAGCTGGT 18213
|||
Qy 1107 euLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrG 1127
|||
Db 18214 ACAAGATGGGAAAGCAGCAGGTGGACCGCCACCCACCATCTCATTTGAAGACCTGATG 18273
|||
Qy 1127 lyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysI 1147
|||
Db 18274 GCTCGTGTGCACTCATCTCGACACCGCTGACCGGTGGAGTCTGGCCAGTACATGTGCT 18333
|||
Qy 1147 leAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLys- 1166
|||
Db 18334 TCGGCGCCAGCGCCGCTGCG-----AATGTCAGTACCTCGGCCCTCAGCCCTTTGTGGAGG 18375
|||
Qy 1167 --GluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluG 1186
|||
Db 18376 TCCTGGTGCAAGTCCCAACACCGTTCGTGAACAAGGTCCGGGCCCTCAGCCCTTTGTGGAGG 18435
|||
Qy 1186 lyHisProValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpL 1206
|||
Db 18436 GAGAGGACGCCAGTTCACCTGCACCATCGAAGCGCGCCCGTACCCCGCAGATCAGGTGGT 18495
|||
Qy 1206 ysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrT 1226
|||
Db 18496 ACAGGACGGGGCCCTGCTGACCACTGGCAACAAGTTCACAGACACTGAGTGAGCGCTCGCA 18555
|||
Qy 1226 hrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrTrpL 1246
|||
Db 18556 GCGGCCCTGCTAGTGTGTGTATCCGGGGCGGCCAGCAAGGAGGACCTGGGGCTCTACGAGT 18615
|||
Qy 1246 euSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIle 1263
|||
Db 18616 GTGAGCTGGTGAACCGCTGGGCTCCGCGGGGCTAGTGGCGAGCTGCGCATT 18668
|||
RESULT 10
US-09-925-300-502
; Sequence 502, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 502
; LENGTH: 3192
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3085)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-502

Alignment Scores:
Pred. No.: 1,59e-20 Length: 3192
Score: 432.50 Matches: 263
Percent Similarity: 34.74% Conservative: 141
Best Local Similarity: 22.61% Mismatches: 437
Query Match: 6.28% Indels: 323
DB: 10 Gaps: 46

US-09-818-990b-2 (1-1320) x US-09-925-300-502 (1-3192)
Qy 98 LysArgLeuSerProAspGlnMetLysHisSerPro-----AsnLeuSerPheGluPro 115
|||
Db 87 AAGACCTATCGGAAGACGACGTGAAGAGATCCAGCCGAGAGATGGATTTCCGTGCC 146
|||
Qy 116 AsnPheCysGlnAspAsnProArgSerProThrSerSerLysGluSerProGlnGluAla 135
|||
Db 147 AACCTG-----CAGCGGCAA 161
|||
Qy 136 LysArgProGlnTyrCysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAla 155
|||
Db 162 GTGAAGCCAAAGACTGTGTCTGAG---GAAGAGAGAGAGAGTGCACAGCCCGCAGCGTC 218
|||
Qy 156 AspPheIleGluGluLeuSerSerLysSerHisSerSerLysArgIleArgPro 175
|||
Db 219 GATTTT-----CGCTCTGTCTGCCCAAGAGGGGACTTCCAAGACCCCGTGCCT 269
|||
Qy 176 Arg-----AlaCysLysAsnHisLysSerLysLeuGluSerGln 188
|||
Db 270 GAGAAGGTGCCACCGCCAAAACCTGCCACCCCGGATTTTCGCTCAGTCTGGTGGCAAG 329
|||
Qy 189 AsnLysValMetGlnGluAsn---SerSerSerPheSerAspLeuSerGluArgArgGlu 207
|||
Db 330 AAGAAATATACAGCAGAGAAATGGCAGCAGCAGTGCAGAGACCTGAATGCCAAGGCAAGTG 389
|||
Qy 208 ArgSerSerValProIle-----ProIle----- 215
|||
Db 390 GAGAGTCCCAAGCCCTGAGCAATGCAGCAATGCAGCCCTTCAGGGCCCTTGAACCCGTGGCAAC 449
|||
Qy 216 -----ProAlaAspThr----- 219
|||
Db 450 GCCAAGCCTGCTGAGACCCCTGAAGCCAATGGCAACGCCAAGCCTGCCGAGACCCCTGAAG 509
|||
Qy 220 -----ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnAlaLys 234
|||
Db 510 CCCATGGCAATGCCAAGCCCTGATGAGAACCTGAAATCCCGCTAGCAAGAGAACTCAAG 569
|||
Qy 235 ArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySer 254
|||
Db 570 ANAGCCTTAAAGATGATGTAAGTGCAGAGAGAGCCCATGCAGGAGCAACATAATGAA 629
|||
Qy 255 SerProSerSerLeuTyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLys 274
|||
```

Db	1587	AACGTGTATGAACCACTGAGCCAAACCCAG-----GAGTCTGAACCTCACA	1631
Qy	580	GlyValLeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHis	599
Db	1632	ACGCTAGGAGAAACCTGAAGAGCCGAGGATGAAGTCGAGGTGTC-----	1679
Qy	600	PheAsnLeuProGluAspAspGlySerGluAlaSerSerGluAlaGlyValValThr	619
Db	1680	-----GAYGATGATGAGAAGAGCCGCGAGGTTGATTACCGGCACACTGACAATCAAT	1730
Qy	620	ThrArgGlnThrArgProAsp-----Ser**GlnGluArgPheAsn-----	633
Db	1731	ACTGAACAAAGAAATGCTGACTTCTACGACATTTGAGGAGAGATTAGGATCTGGGAAATTT	1790
Qy	634	GlyGlnAlaThrLysThrProGlu-----	641
Db	1791	GGACAGGCTCTTCGACTTGTAGAAAGAAAACCTCGAAAAGTCTGGCGCAGGGAAGTTCTTC	1850
Qy	642	ProSerPheProValLysGluProProProVal-----	652
Db	1851	AAGGCATATTTCAGCAAAAGAGAAAGAGATAATCCGCGAGGAGATTAGCATCATGAACATGC	1910
Qy	653	LeuAlaLysProLysLeu-AspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLe	672
Db	1911	CTCCACACCCCTTAAGCTGGTCCAGTGTGTGGATG-----	1944
Qy	672	uGluGlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe**We	692
Db	1945	-----CCTTTGAGAAAAGGCCACATCGTCA-----TGGTCTCTGGAGATC	1985
Qy	692	tThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysAl	712
Db	1986	GTGTCAGGAGGGAGCTGTTTGGCGCATCATTTGACGAGCACTTTGAGCTGACGCGAGCGT	2045
Qy	712	aProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnTh	732
Db	2046	GAGTSCATCAAGTACATGCGCGAGATCTCGGAGGGATGG-----AGTACATCCAC	2096
Qy	732	rThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSerSerThrProG1	752
Db	2097	AAGCAGGCGCTGTCACCTGGACCTCAAGC-----CGGAGNACATCATGTCTGTC	2147
Qy	752	nThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerValGlnTh	772
Db	2148	AACAAGACGGGCACCAAGGATCAAGTCATCGACTTTGGTC-----	2187
Qy	772	rLysSerProGlyGly-----LeuSerIleGlnAs	782
Db	2188	-----TGGCCAGGAGGCTGGAGAACGCGGGTCTCTCAAGGTCTCTTTTGGCACCCAGAA	2243
Qy	782	nGluProLeuProProGlyProThr---GluProThrProProThrPheSerI11	801
Db	2244	TTTGTGGCTCTGAAGTGATCACTATGAGCCCATCGGCTAGCCACACACATGTGGA---	2301
Qy	801	eProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThrSe	821
Db	2302	-GCATCGGGGFCATCTCTACATCCCTAGTCAGTGGCC---TTTCCCCCTTCATGGGAGAC	2357
Qy	821	rArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIlePr	841
Db	2358	AACGATACACGAAACCT-----TGGCCAAAGCTTACCTCAGCCACC	2399
Qy	841	oProThrAsnAlaMet**LeuProArgSerAlaProSerMetPro-----	856
Db	2397	TGGGACTTCGACGACGAGGCAATTGCGATGAGATCTCGACGATGCCAAGGATTTCATCAGC	2456
Qy	857	-----SerGlnGlyLe	860
Db	2457	AATCTGCTGAAGAAGATATGAAAAACCGCGCTGGACTGCACGCATGCTTTTCAGCATCCAT	2516
Qy	860	uAlaLys-LysAsnThrLysSer-----ProGlnProValAsnAspAsnIleArg-	877
Db	2517	GGCTAATGAAGATACCAGAAACATGGAGGCCAAGAAACTCTCCAAGGACCGGATGAAGA	2570

```
Qy 878 -----GluThrLysAsnAlaValIleArgAspLeuGly- 888
Db 2577 AGTACATGGCAAGAAATGGCAGAAACGGCAATGCTGTG---AGAGCCATTGGAA 2633
Qy 889 -----LysLysIleThrPheSerA 895
Db 2634 GACTGTCTCTATGGCAATGATCTCAGGGCTCAGTGGCAGGAAATCTCTCAACAGGGTCAC 2693
Qy 895 spValArgProAsnGlnGlnGluThrLysIleSerSerPheGluGlnArgLeuMetAsnG 915
Db 2694 CAACAGCCCTCAATCGAGAA---AAACTAGATCTGAAGAAGATGTGCCAAGCTT 2750
Qy 915 luIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisA 935
Db 2751 TCCTTGAGGCTGTGCTGAGGAAAGCCT-----CAT- 2782
Qy 935 spGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheA 955
Db 2783 -----GTAAACCCCTATTCTCTAAGACCATTCGCGATTAG 2819
Qy 955 rgValThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProL 975
Db 2820 AAGTTGTGGGGAAGTCTCTGCTAGATTGACTGCAAGATTGAAGGATACCACGACCCCG 2879
Qy 975 ysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMeta 995
Db 2880 AGGTTGTCTGTTCAAGATGACCCAGTCAATC---AGGAGTCCCGCCACTTCAGATAG 2936
Qy 995 rgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspG 1015
Db 2937 ACTACGATGAGGAGCGGGAACCTGCTCTTTAAATTATTAGTGATGTTTGGCGGGATGACGATG 2996
Qy 1015 lyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuM 1035
Db 2997 CCAAGTACACCTGCAAGGCTCTCAACAGCTCTGGAGAAAGCCACCTGCACAGCAGCTCA 3056
Qy 1035 etValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyA 1055
Db 3057 TTGTGGAAACGATG-----GAGGAAGGTGAAGNGA-AGGGGAA 3094
Qy 1055 rgSer---ArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProH 1074
Db 3095 GAGGAAGAAGTAGTGAACAAAGCCAGAGAAAGCAGTTTCTTAAGTCATATTAAAGGACT 3154
Qy 1074 lsPhe 1075
Db 3155 ATTTC 3159

RESULT 11
US-09-905-129-1
; Sequence 1, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905.129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8883
; TYPE: DNA
```

```
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(8883)
; OTHER INFORMATION: 'n' can be any nucleotide 'a', 'c', 'g' or 't'.
US-09-905-129-1

Alignment Scores:
Pred. No.: 6,73e-19 Length: 8883
Score: 416.50 Matches: 354
Percent Similarity: 30.73% Conservative: 216
Best Local Similarity: 19.08% Mismatches: 552
Query Match: 6.04% Indels: 733
DB: 10 Gaps: 83

US-09-818-990B-2 (1-1320) x US-09-905-129-1 (1-8883)

Qy 2 GlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu 18
Db 1460 CAGGAGGACAATGGATCTGCCCTCCACCTCAAGATTTTCATAGAACCCCTTTGGCTCC 1519
Qy 19 SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla 35
Db 1520 TTGTCTTTGAACATGACGACANANNNTCTGGAATAAAGGCCGACATGGTCTGTAGTATCCAA 1579
Qy 36 GluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 55
Db 1580 AAGCCATCAAGGACA-----TCACCACTGCATTCACTGAA-----TCC 1615
Qy 56 GlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSer 75
Db 1616 ---GAAATGACTACATCATCTAAATCGCTCATTT-----TCC 1651
Qy 76 ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln 95
Db 1652 ACAATCTGTG---TGCAGTGTAGATTAAATCACATCCAGCCAGTGTGGCAACTTCTG 1708
Qy 96 AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro 115
Db 1709 GCTTATACAGTGACTCTCTCTGATAGAAAGGAGCCAGCTTACC-----TACC 1759
Qy 116 AsnPheCysGlnAspAsnProArgSerProThrSerSerLysGluSerProGlnGluAla 135
Db 1760 -----GAGACTCTTCACTGTCTTCTAGATATAAACAGGTGGCT 1798
Qy 136 LysArgProGlnTyrCysSerGluThr----- 144
Db 1799 CTTAGGCTGAAGACATTTTACCAGCATAGAGCTGATGTGACAGCAGACCCCTTTTGG 1858
Qy 145 ---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluLeu 161
Db 1859 TTCCAACAAGAAAAAATGTCTTCAGCTGAACAGAACTGCCACACACTTAGCACATTA 1918
Qy 162 SerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 1919 CAGATCCAGTTTTCACCTGATGCTCAAAATCGCTTTA---CCAAGGGCG---GAGATGAGA 1972
Qy 182 LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp 201
Db 1973 GCGGAGAGACTCAAAATGGACCATGATCTGTATGATGAACAAT----- 2014
Qy 202 LeuSerGluArgArgGluArgSer-----SerValPro 212
Db 2015 -----CCCAAACTGGAACGCACTGCTCTGTTGGCGSCACTATTGCCCTCAGCTGTCCA 2068
Qy 213 -----IlePro 214
Db 2069 GGCAAGCGCACCTTCACCTCACTTGAATGGCTTCTAGCTGATGGAGTAAAGTGAGA 2128
Qy 215 IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln 230
Db 2129 GCCCTTACGTTGCGGAGGATGGCGGAATCTTAATAGACAAAAATGGGAAGTTGGAACGTG 2188
```


D	b	4268	GCAAGCTTCCAAAATAGCTCCTCTTTTACCCACAGGTCAGAGTTCCCCTCAGATTCT	4320
O	y	693	ThrValLeuAsnSerAsnAlaProProAlaValThrThrSer ::::: ::: :::	706
D	b	4328	ACAACACTCTGTGACAGTCCGCCACCAGCTGTCTACAACAANTGGTGCGCACTCAGAAC	4387
O	y	707	*****LysGlnValLysAlaProSerSerGlnThrPhe ::: :::	718
D	b	4388	AAGGCACACTGAAGTAGTATCAGGTGCCAAGAGTCTCTCAGCAGGGAAGAACGACCCCTTC	4447
O	y	719	SerLeuAlaArgProLysTyrrPhePheProSerThr-----AsnThrThr :::	733
D	b	4448	ACCACCTCCTCTCCA-----GTGCTTCCTAGCACCATTAAGCAAGAGATCTTAATACATTA	4501
O	y	734	----AlaAlaThrValAlaProSer-----SerSerProValPheThrLeuSerSerThr ::: :::::	750
D	b	4502	AACTTCTGTCAACAGGAACCCOCACAGTGCAGAGTCTCTACTGCTACTGCACTGTGTCATT	4561
O	y	751	ProGlnThrIleGlnArgThrValSerLysGluSer----- :::::	762
D	b	4562	ATGCTCAAACCCCAACGAACAGATCCAAAGAAAGCAACCAAAATAAAGGGCGCTCGG	4621
O	y	762	----- -----	762
D	b	4622	AAGAACAAGAAACACGCAAAACACACCCCCAGCAGGTTTTCTGGCTATAGTGCACTACTCA	4681
O	y	763	-----LeuLeuValSerHisProSerValGln----- 	771
D	b	4682	GCTTAAACAACAGCTGATACCCCTTTGGCTTTCAGTCAATCCCCACGACAAGATGATGGT	4741
O	y	771	----- -----	771
D	b	4742	GGAAATGAAGTCAGTGTCTTATCACTCAACAACCTCTCTCTGCGCCATAAAGTGAACCTG	4801
O	y	772	-----Thr -----	772
D	b	4802	TTTGAGAAGTACACCCAGACATTTGGGAAATACAACAGCTTTGGAAACAACGTTGTTGAGC	4861
O	y	773	LysSerProGlyGlyLeuSerIleGlnAsn-----GluProLeu----- :::::	785
D	b	4862	AAATCAGGAGAGTACCAGTGAAGAGAGCTCAGACACACACACCACTCCTCAGC	4921
O	y	786	-----ProProGlyProThrGluProProProPheThrPheSerIlePro :::::	802
D	b	4922	AGTGGCGCGCCCGCAGTCCACT--CCTTCCACACCTCCTTTACTAAGGGTGTGGTT	4978
O	y	803	SerGlyAsnGln-----PheGlnProArgCysValSerProIleProValSer ::: ::::: :::	818
D	b	4979	ACACACAGCAAGATCATATCAGCTTTCCAGATGAGCTCAAAATAGATGTTGCCATATAT	5038
O	y	819	ProThrSerArg-----IleGlnAsnProValAla----- :::	828
D	b	5039	GAATCTTCAAGGCACAATACAGATCTGCAGAACCCCTCAGCAGAGGCTAGCCCAATCCT	5098
O	y	829	-----PheLeuSerSerValLeuPro :::::	835
D	b	5099	GAGATCATACTGGAACCACTGACTCTCCTCTAATCTGTTTCCACTCCACTCTGTGCCA	5158
O	y	836	SerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMet :::	855
D	b	5159	GCATAAGGGTAGATAAACCAACAGAATCTTAATGAAGCCC----- -----	5200
O	y	856	ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAsn 	875
D	b	5201	-----TCTCCCTGGCCA----- -----	5212
O	y	876	IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysIleThrPheSerAsp -----	895
D	b	5212	----- -----	5212
O	y	896	ValArgProAsnGlnGlnGluTyrrLysIleSerSerPheGluGlnArgLeuMetAsnGlu ::: :::::	915
D	b	5213	-----CAACACAAATATCAGCTCAAGTCACTACTCCGAACCACTTGAGAAGGCG ::: :::::	5260

Qy	916	IleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAsp	935
Db	5261	AAAAGGCGCAGCATGAAGCATGTCCCC	5293
Qy	936	GluIleProThrGlyLysCysIleAla	944
Db	5294	AGCCTTCAGAGCCAGCACTCATGCTCTACACTGGATATACAGAAGCATGCGAAGAAAG	5353
Qy	945	ProIlePheAspLysArg	950
Db	5354	AGTGTGTTTATAGAAGAACTGGTCAAAACCAACTTCCAAACATCTGCCTTACGTCTCT	5413
Qy	951	LeuLysHisPheArgValThrGluGlySerProValThrPheThr	965
Db	5414	CTACCTAAGACTCTATTGAAAGAGCCAAAGAAATTAATTTGGAGGAAAGCGTGCAGAGCTTTACA	5473
Qy	966	CysLysIleValGlyIleProValProLys	975
Db	5474	GTTCACAGCTAATTCAGACGTTTTTCTTCCTGTGAGGCTGTGGAGACCACTGCCCATC	5533
Qy	976	ValTyrTrpPheLys	993
Db	5534	ATCCACTGGACAGAGTTTCATCAGGANTTGAATATATCCAAAGGCACACAGAAAGCGCG	5593
Qy	994	MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp	1013
Db	5594	TTCACAGCTGCTCCCAATGACAC	5647
Qy	1014	AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis	1033
Db	5648	CGTGGACAGTACCTGTGCTGCTGATTAATCCACTGGCGTAGACCATTTTCATGTCTCT	5707
Qy	1034	LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyClnSerHisArg	1053
Db	5708	TTGTCTGTGTTTTTTACCG	5728
Qy	1054	GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPro	1073
Db	5729	GCAGGATTTGGACACACATGTCAG	5755
Qy	1074	HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAsp	1093
Db	5756	GAGATTCACAGTTCAC	5794
Qy	1094	CysLysValSerGlyLeuProProGluLeuThrTrpLeuAsnGlyClnProVal	1113
Db	5795	TGCAGAGTGAGGTATGCCGAGCCCTACGGTTTCTGGATACCTGCAACCAACCGGTG	5854
Qy	1114	LeuProAsp	1130
Db	5855	GTCTCAGAAACGGCCAAAGGAAGCAAGGCTGGGTAAACACCTGATGGA	5908
Qy	1131	LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn	1150
Db	5909	TTGATCATCTATATCTGAGTCTTTATGATCGTGGTTTTTACAGTGTGGCCGAGCAAC	5968
Qy	1151	LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys	1170
Db	5969	CCATCTGCCGAGGATTCACCTGTTGGTTAAGATACAAGTCATCACA	6016
Qy	1171	AlaProValIleLeuGluLysLeuGlnAsnCys	1188
Db	6017	CCCCCTGCTCATATAGAGCAAAAGAGGCAAGCCATCGTTTGGGGTTT	6073
Qy	1189	ValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAsp	1208
Db	6074	TTGAAACTGCCCTGCACCTGCAAAGAGAACTCCCGACCTAGTGTTCATCGGTCCTTAT	6133
Qy	1209	AsnGluThr	1224
Db	6134	GATGGGACTCAACTAAACCATTCAGTTCAGTCACTCCAGATTTTTCTCTGATCCAAAT	6193

Qy	763	-----IeuLeuValSerHisProSerValGln-	771
Db	4682	GCTCTAACACAGCTGATACCCCTTGGCTTTTCAGTCATATCCCCACGACAGATGATGGT	4741
Qy	771	-----	771
Db	4742	GGAATGTAAGTGCAGTTGCTTATCCTCAACAACCTCTCTTCGGCCATAACTGAAC TG	4801
Qy	772	-----Thr	772
Db	4802	TTTGAGAAGTACACCCAGACCTTTGGAAATACACAGCTTTGGAAACAACGTTGTTGAGC	4861
Qy	773	LysSerProGlyLeuSerIleGlnAsn-----GluProLeu-----	795
Db	4862	AAATCAGGAGAGTAGTACCACAGTGAAGAGAGCCTCAGACACACACACCACTCCTCAGC	4921
Qy	786	-----ProProGlyProThrGluProThrProProProPheThrPheSerIlePro	802
Db	4922	AGTGGGGCGCCCGACTGCGCACT--CCTGCCCACTCTCTTTACTAAGGTGGTGGTT	4978
Qy	803	SerGlyAsnGln-----PheGlnProArgCysValSerProIleProValSer	818
Db	4979	ACAGACAGAAAAGTCACATCAGCTTCCAGATGACGTCAAATAGAGTGGTCACCATATAT	5038
Qy	819	ProThrSerArg-----IleGlnAsnProValAla-----	828
Db	5039	GAATCTTCAAGGCACAATACAGATCTGCAGCAACCTCAGCAGAGGTAGCCCCAATCCT	5098
Qy	829	-----PheLeuSerSerValLeuPro	835
Db	5099	GAGATCATAACTGGAACCACTGACTCTCCCTCTAATCTGTTTCCATCCACTCTGTGCCA	5158
Qy	836	SerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMet	855
Db	5159	GCATTAAGGCTAGATAAACACACAGAATTCTAAATGGAAGCCC-----	5200
Qy	856	ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAspAsn	875
Db	5201	-----TCTCCTCGGCCA-----	5212
Qy	876	IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAsp	895
Db	5212	-----	5212
Qy	896	ValArgProAsnGlnGlnGluTyrIleSerSerPheGluGlnArgLeuMetAsnGlu	915
Db	5213	-----GAACACAAATATCAGTCAAGTCATCTCCGAAACCATTTGAGAAGGC	5260
Qy	916	IleGluPheArgLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAsp	935
Db	5261	AAAAGCCAGCAGTAGCATGTCCCC-----CACCTC	5293
Qy	936	GluIleProThrGlyLysCysIleAla-----	944
Db	5294	AGCCTCCAGAGGCCAGCACTCATGCTCACCCTGGAATACACAGAAGCATGCAGAAAG	5353
Qy	945	ProIlePheAspLysArg-----	950
Db	5354	AGTGTGTTTGTATGAAGAACTGGTCAAAACCACTTCCAAACATCTGCCTTAGCTCTCT	5413
Qy	951	-----LeuLysHisPheArgValThrGluGlySerProValThrPheThr	965
Db	5414	CTACCTAAGACTCTATTGAAAAGCCAGATTAATTGGAGGAAGGCTGCAAGCTTTACA	5473
Qy	966	-----CysLysIleValGlyIleProValProLys	975
Db	5474	GTTCACAGCTAATTCAGACGTTTTTCTTCCTTGTGAGGCTGTTGGAGACCCACTGCCCATC	5533
Qy	976	ValTyrTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys	993
Db	5534	ATCCACTGGACCAAGCTTTCATCAGAGANTTGAATATATCCCAAGGCACACAGAAAGCCGG	5593

[illegible]

RESULT 13

US-09-905-129-7
; Sequence 7, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 8883
; TYPE: DNA
; ORGANISM: Rattus sp.
; NAME/KEY: misc_feature
; LOCATION: (1)..(8916)
; OTHER INFORMATION: 'n' can be any nucleotide 'a', 'c', 'g' or 't'.
US-09-905-129-7
Alignment Scores:
Pred. No.: 6,73e-19 Length: 8883
Score: 416.50 Matches: 354
Percent Similarity: 30.73% Conservative: 216
Best Local Similarity: 19.08% Mismatches: 552
Query Match: 6.04% Indels: 733
DB: 83
US-09-818-990b-2 (1-1320) x US-09-905-129-7 (1-8883)
Qy 2 GlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu 18
Db 1460 CAGGAGGACAAATGATGCTCCCTCCACCTCAAGATTCATAGAACCCCTTTGGCTCC 1519
Qy 19 SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla 35
Db 1520 TTGCTTTGACATGACACANNTTCTGGAATTAAGCCGACATGGCTGTAGTATCCAA 1579
Qy 36 GluProSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 55
Db 1580 AAGCATCAAGGACA-----TCACCACTGCATTCACTGAA----- 1615
Qy 56 GlyGlnAspSerLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSer 75
Db 1616 ---GAAATGACTACATCATGCTAAATCGTCATTT-----TCC 1651
Qy 76 ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln 95
Db 1652 ACAATCTTGTG---TGCAGTGTAGATTATATACATCCAGCCAGTGTGCACTCTCTG 1708
Qy 96 AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro 115
Db 1709 GCTTTATACAGTACTCTCCTCTGATGATAAGAAAGGAGGAGCCAGCTTACC----- 1759
Qy 116 AsnPheCysGlnAspAsnProArgSerProThrSerSerLysGluSerProGlnGluAla 135
Db 1760 -----GAGACTCTTCACTGCTCTTAGATATAAACAGGTGGCT 1798
Qy 136 LysArgProGlnTyrCysSerGluThr----- 144
Db 1799 CTAGGCTGAGACATTTTACCAGCATAGAGCTGTGTCAGAGCAGACCCCTTTTGG 1858
Qy 145 ---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluLeu 161
Db 1859 TTCCAACAAGAAAAAATTTGCTTGCAGCTGAACAGAACTGCCACCACTTAGCACATTA 1918
Qy 162 SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 1919 CAGATCCAGTTTCCATCATGATGATGATGATGATGATGATGATGATGATGATGATG 1972
Qy 182 LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp 201
Db 1973 GCGGAGAGACTCAATGGACCATGATGATGATGATGATGATGATGATGATGATGATG 2014
Qy 202 LeuSerGluArgArgGluArgSer-----SerValPro 212
Db 2015 -----CCCAAACTGGAACGCACTGCTGCTGCTGCGGCACTATTGCGCTGAGTCCA 2068
Qy 213 -----IlePro 214
Db 2069 GGCAAGCGCACCCTTCACTCACTTGAATGGCTTCTAGCTGATGGAGTAAGTGA 2128
Qy 215 IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln 230
Db 2129 GCCCTTACGTTAGGAGGATGGCGAATCTAATAGACAAATGGGAAGTTGGAAGTG 2188
Qy 231 GlnGluAlaLysArgArgGluAla-----GluGlnAlaLysSerGluAlaAlaGly 247
Db 2189 CAGATGGCTGACAGCTTTGATGACAGGCTTTTACCACCTGCATAAGCACAATGATG 2248
Qy 248 GlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGlu-----Glu 263
Db 2249 GCGGATGTT-----CTCACATACAGGATAACTGTGTTAGTGGATT 2284
Qy 264 ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly 283
Db 2285 CCCTATGGAGAAACACACATGACAGTGGAGTC-----CAGCACAGAGTGGTTACGGGT 2338
Qy 284 ThrArgValGlnLeuAspCysIleValValGlyIleProProGlnValArgTyrTyr 303
Db 2339 GAGACGCTCGACCTTCCATGCTTCCACGGGTGTTCCAGATGCTTCTATAGTGGATT 2398
Qy 304 CysGluGlyLysGluLeuGluAsnSerPro-----AspIleHisIleValGlnAlaGly 321
Db 2399 CTTCCAGGAAACACTGTGTTCTCAGCCATCAAGACAGAGCAAAATCTTAAACAATGGG 2458
Qy 322 AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys 341
Db 2459 ACCTTAAGATATTACAGTTACG-----CCAAAAGATCAAGTCAATCAATGT 2509
Qy 342 PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIle----- 358
Db 2510 GTGGCTGCCAACCCATCAGGGCGGCACTTTTCCAGTTTAAAGTTTCAAGTTCAAAAGAAA 2569
Qy 359 ---GluGlyValSerSerSerAspSerGlu-----GlyAspProAsn 371
Db 2570 GGCCAAAGGATGGTTGACATGACAGGAGGAGCAGGTGGATCGACTTGGAGAACCAAC 2629
Qy 372 LysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerProThrThrSer 391
Db 2630 TCCAGTGTTCCTTAAAGCAGCCAGCATCTTTGAAACTCTCTGCATCAGCTTACAGGG 2689
Qy 392 AlaValIleProProAlaValPro----- 399
Db 2690 TCAGAGGCTGGAAGAAACAGTCTCCGGTGTACATAGGAAGAACAAACATAGAGACTTAATA 2749
Qy 399 ----- 399
Db 2750 CATCGCGCGCTGGGATTCACGCTCCGGGATTCAGGGAGCATAGAGCAGCTCCCT 2809
Qy 400 ---GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGln--- 417
Db 2810 CTCTCTGCTGGAGAAATTGACCCGCAAGCTTTCAGAGCAGCACTTCTAGAAAAAGCCAAAAG 2869
Qy 418 ---SerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle-----IleAlaAlaPro 435
Db 418 ---SerProThrAsnTyrLeuGlnGlyLysProIle-----IleAlaAlaPro 435

Db 2870 AATTCTGTGCCAAAAAGCAGAAAAATACCACAGATAAGCCAGTGCCTGCTTCCC 2929
Qy 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSer-----GluGlyGln 450
Db 2930 CTCGTGAACACTGACGAGGAGAAAGATGCGCTCTGGCATGATTCCTCAGATGAAGAA 2989
Qy 451 LeuValValPheGluCysArgValLysGlyAlaPro-----SerProLysValGlu--- 467
Db 2990 TTCATGCTTCTGAAACATAAGCTTCTGGTCCCGACGAGAGGTCCACCAACTGCTGACTCT 3049
Qy 468 -----TrpTyrArgGluGlyThrLeuIle----- 475
Db 3050 GGACCACTAAATCATGTTTTATGACGAGTATAGCTTCTGGCACAGAGTCTCAACTGTG 3109
Qy 476 -----GluAspSerProaspPheArg----- 482
Db 3110 AATCCACAACACTACAATCTGACACCTCTCTGATTTCAAAATTATTATTAGTGTAAACAAC 3169
Qy 483 -----IleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThr 499
Db 3170 GGTACAGCTGTGACAAAGATGATGAACCCATCCATAGCAAGCAAAATAGAAATACAAACC 3229
Qy 500 -----LeuValIleAlaGluValPheAlaGlu-----AspSer----- 510
Db 3230 AACCAAAACCCAAATCATATTATCTTCCATCAGTAGCTGAAATTCGAGATTCTGCTCAGGCA 3289
Qy 510 ----- 510
Db 3290 GGAAGACATCTTCCCAAGTGCACACCCCTGTAAACAGGGGAAACATGGCTACCTATGGC 3349
Qy 511 -----GlyCysPheThrCysThrAlaSer----- 518
Db 3350 CATACCAACACATATAGTAGCTTTTACCAGCAACCCAGTACAGTCTTCCAGCCCAATAAAT 3409
Qy 519 -----AsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGlu 536
Db 3410 CCAACAGAAAGTTATGGACCTCAGATACCTATTACAGAGAGTCAGACACCTAGCAGTAGT 3469
Qy 537 AspLeuSerAsnAngly-----SerLeuHisSerAlaAsnSerThr 550
Db 3470 GACATCTCTCTACACTACTGACAGCCCTAGCTTCTCCAGTACCCCTTCAGGTTACAC 3529
Qy 551 ThrAsnLeuAlaIleGluProGlnPro----- 560
Db 3530 ACCACTGCCCTGCTTTATTTCATATCCCTAGAACACACATACAGGTAACTTCCCTTG 3589
Qy 561 -----SerProProHis 564
Db 3590 TCCAGGCACCTTGGGAGAGAGAGGACAAATTTGGAGCAGAGGGAGAGTTAAAAACCCACAT 3649
Qy 565 SerGluPro----- 567
Db 3650 AGAACCCAGTTCTCCGACGGCATAGACACAGGACTGTGAGGCCGACCAATCAAGGGACCT 3709
Qy 568 -----ProSerValGluGlnPro----- 573
Db 3710 GCTAACAAAAATGTGAGCCAAAGTTCCAGCCACAGAGTACCCTGGGATGTGCCACACATGT 3769
Qy 573 ----- 573
Db 3770 CTTTCCGACAGGGGCTCACAGTGGCTACTGCAGCACTGTCCAGTTCCAAGTTTCACTCCAC 3829
Qy 574 -----ProLysProLysLeuGluGlyValLeuVal----- 583
Db 3830 AGTCCCTCCCCAAACTAATAATGTTGGGGTCATAGCAGAAAGAGTCTACCACTGTGGTC 3889
Qy 583 ----- 583
Db 3890 AAGAAACCACTGTTACTATTAAAGCAACAACAAATGTAGATATTGAGATAATAACAAC 3949
Qy 584 -----AsnHisAsnGluProArgSerSerArgile 594
Db 3950 ACTACAAAATATTCCGGAGGGGAAAGTAACCACTGATTCTCAGGAGGAAGCATGACT 4009

Qy 595 GlyLeuArgValHisPheAsnLeu-----ProGluaspaspLysGlySerGluAla 611
Db 4010 TCTGCTCCACATCTGTATCCCTGGGAAATCTCTCTAGACAAATAGTGT----- 4060
Qy 612 SerSerGluAlaGlyValValThrArgGlnThrArgProAspSer***GlnGluArg 631
Db 4061 -----CACCTGAGCATGCCTGGGACCACCAAACTGGAAAGATTTCAGTGGAA----- 4108
Qy 632 PheAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProPro--- 650
Db 4109 -----ACAACACCACTTCCCCAGC-----CCCTCAGCAGCACCTCAATA 4147
Qy 651 ProValLeuAlaLysProLysLeuAspSerThrGlnLeuGlnLeu-----HisAsn 668
Db 4148 CCAACAGCACAAAATTTCTCAAGAGGAAACTCCCTTGCACCAAGATCTTTGTAATAAAC 4207
Qy 669 Gln-----ValLeuLeuGluGlnHisGlnLeu-----GlnAsnProPro 681
Db 4208 CAGAAGAAGGAGGGGATGTTAAAGAATCCATATCAATTCGGTTTACAAAAGAACCCAGCC 4267
Qy 682 ProSerSerProLys-----GluPheProPhe***Met 692
Db 4268 GCAAGAGCTTCCCAAAATAGCTCTCTTTTACCCACAGTCCAGAGTTCCTCCCTCAGATTCT 4327
Qy 693 ThrValLeuAsnSerAsnAlaProAlaValThrThrSer----- 706
Db 4328 ACAACTCTTGGACAAGTCGCCACAGCTCTGTCTACAACAATGGCTGCCACTCAGAAC 4387
Qy 707 -----***LysGlnValLysAlaProSerSerGlnThrPhe 718
Db 4388 AAGGCACCTGAAGTAGTATCAGGTGCCAGAGTCTCTCAGCAGGGAAGAACGCCCTTC 4447
Qy 719 SerLeuAlaArgProLysTyrPhePheProSerThr-----AsnThrThr 733
Db 4448 ACCAACTCCTCTCCA-----GTGCTTCTCAGCACCATAAGCAAGAGATCTTAATACATTA 4501
Qy 734 -----AlaAlaThrValAlaProSer-----SerSerProValPheThrLeuSerSerThr 750
Db 4502 AACTTCTTGCACGGAAACCCCCACAGTGACAAGTCTACTGTCTACTGTGATCTGTCAAT 4561
Qy 751 ProGlnThrIleGlnArgThrValSerLysGluSer----- 762
Db 4562 ATGCTGAACCCCAACGACGAAGATCCAAAGAGCAAAAGACCAATAAAGGGCCCTCG 4621
Qy 762 ----- 762
Db 4622 AAGAACAGAAACAAACGCAACACACCCACCCAGCAGGTTTCTGGCTATAGTGCATACTCA 4681
Qy 763 -----LeuLeuValSerHisProSerValGln----- 771
Db 4682 GCTTAACACAGCTGATACCCCTTGGCTTTCAGTCATCTCCCAACGACAGATGATGGT 4741
Qy 771 ----- 771
Db 4742 GGAATGTAGTGCAGTGTCTTATCACTCAACAACCTCTCTTCTGGCCATAACTGAACGT 4801
Qy 772 -----Thr 772
Db 4802 TTTGAGAAGTACACCCAGACTTTTGGGAAATACACAGCTTTGGAACAACACCTGTTGAGC 4861
Qy 773 LysSerProGlyGlyLeuSerIleGlnAsn-----GluProLeu----- 785
Db 4862 AAATCAGCAGGAGATACACAGTACCCAGTGAAGAGCCCTCAGACACACCCACCTCCTCAGC 4921
Qy 786 -----ProProGlyProThrGluProThrProProPheThrPheSerIlePro 802
Db 4922 AGTGGGGCGCCCGCAGTCCCACT---CCTTCCCACTCCTTTTACTAAGGGTGTGGTT 4978
Qy 803 SerGlyAsnGln-----PheGlnProArgCysValSerProIleProValSer 818
Db 4979 ACAGACAGCAAAAGTCACATCAGCTTTCAGATGACGTCAAATAGATGCTGCATATAT 5038

```
QY 819 ProThrSerArg-----IleGlnAsnProValAla----- 828
DB 5039 GAATCTTCAAGGCACAATACAGATCTGCAGCAACCTTCAGCAGAGGCTAGGCCCAATCCT 5098
QY 829 -----PheLeuSerSerValLeuPro 835
DB 5099 GAGATCAATACTGGAAACCAGCTCTCCCTCTAATCTGTTCATCCACCTCTGTGTGCCA 5158
QY 836 SerLeuProAlaIleProProThrAsnAlaMet**LeuProArgSerAlaProSerMet 855
DB 5159 GCACCTAAGGGTAGATAAACCCAGCAATCTTAATGGAAGCCC----- 5200
QY 856 ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAsn 875
DB 5201 -----TCTCCCTGGCCA----- 5212
QY 876 IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAsp 895
DB 5212 ----- 5212
QY 896 ValArgProAsnGlnGluThrLysIleSerSerPheGluGlnArgLeuMetAsnGlu 915
DB 5213 -----GACACAATATCAGCTCAAGTCATCTACTCGAACCATTGAGAGGGC 5260
QY 916 IleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAsp 935
DB 5261 AAAGGCCAGCAGTAAGCATGTCCCCC-----CACCTC 5293
QY 936 GluIleProThrGlyLysCysIleAla----- 944
DB 5294 AGCCTTCAGAGGCCAGCAGCTCATGCTCATCTGGAATACACAGCATGCAGAAAG 5353
QY 945 ProIlePheAspLysArg----- 950
DB 5354 ACTGTTTGTGATAAGAAACCTGGTCAAAACCCAACTTCCAAACATCTCCCTTACGTCCT 5413
QY 951 -----LeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
DB 5414 CTACCTAAGACTCTATTGAAAAGCCAAAGATAATTGGCAAGAGGCTGCAAGCTTTACA 5473
QY 966 -----CysLysIleValGlyIleProValProLys 975
DB 5474 GTTCAGCTAATTCAGACGTTTCTCTCTGTGTGAGGCTGTGGAGACCCACTGCCCATC 5533
QY 976 ValTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
DB 5534 ATCCACTGGACAGAGTTTCATCAGGANTTGAATATCCCAAGGGACACAGAAAGCCGG 5593
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
DB 5594 TTCCACGTTGTTCCCAATGGCACC-----TTGTCCATCCAGAGGGTCAGTATTACAGGAC 5647
QY 1014 AspGlyAsnThrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033
DB 5648 CGTGACAGTACTGTGCTGCTGCAATTAATCCACTGGCGGTAGACCAATTTTCATGTCCT 5707
QY 1034 LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArg 1053
DB 5708 TTGTCGTGTTTTTACCG----- 5728
QY 1054 GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArgPro 1073
DB 5729 -----GCAAGGATTTTGGACAGACATGTCAAG----- 5755
QY 1074 HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuLysArgLeuAsp 1093
DB 5756 -----GAGATCACAGTTTCATTTTGAAGGTACTGTGGAACATAAG 5794
QY 1094 CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVal 1113
DB 5795 TGCAGGTGGAGGTATCCCGAGGCTACGGTTCTCTGGATCTTGCACAAACCAACCGTG 5854
QY 1114 LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
```

```
DB 5855 GTCACAAAAGGCCAAGGAAGCAGAAAGCTCTGGGTAAACACCTGATGGA-----ACA 5908
QY 1131 LeuLeuIleLeuAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThrAsn 1150
DB 5909 TTGATCATCTATATCTAGCTTTTATGCTGTTTATGAGTGTGTGGCCAGCAAC 5968
QY 1151 LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys 1170
DB 5969 CCATCTGCCCAGGATTCACCTGTTGTTAAGATACAGTCATCACA-----GCT 6016
QY 1171 AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro 1188
DB 6017 CCCCTGTCATATTAGAGCAAAAGAGGCAAGCCATCGTTGGGTT---TTAGGTGAAGT 6073
QY 1189 ValArgLeuGluCysArgValIleGlyMetProProValPheTrpTrpLysLysAsp 1208
DB 6074 TTGAACCTGCCCTGCACCTGCAAAAGAACTCCCGAGCTAGTGTTCACCTGGTCTTTAT 6133
QY 1209 AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp 1224
DB 6134 GATGGACTGAACATAAACCAATTCAGTTGACTCATTCACAGATTTTCTGTATCCAAAT 6193
QY 1225 ThrThrGlyTyAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyr 1244
DB 6194 GGAAC-----CTGATATAAGAAAGCATCGCTCTTCAGTGAGGGCAGCTTAT 6241
QY 1245 ThrLeuSerAlaLysAsnGluAlaGly----- 1253
DB 6242 GAGTGCAATTCACACAGCTCCTCAGGCTCAGAGAGAGGGTAGTGATTTCTTACTGTGAA 6301
QY 1253 ----- 1253
DB 6302 GAGGAGAGACAATCCCGAGGATAGAAACTGCCTCTCAGAATGAGCTGAGGTGAATTG 6361
QY 1254 -----IleValSerCysThrAlaArgLeuAsp----- 1262
DB 6362 GGTGAGAAATTAATACTGCACTGCTCAGCTACTGGGATCCAAAGCCCTAGAAATAATCTGG 6421
QY 1263 -----IleTyAlaGlnTrpHisGlnIleProProMetSer 1276
DB 6422 AGGTGTCATCCAGGCTGTCATCGACCAAGTGG---CACAGAATGGGAGCCGAATCCAC 6478
QY 1277 ValArgProSerGlySerArg---TyrGlySerLeuThrSerLys 1290
DB 6479 GTCCTACCAAAATGATCCTTGTGTTGGTTCAGTGACGGAAAAA 6523
RESULT 14
US-09-991-630-1
; Sequence 1, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991.630
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8883
; TYPE: DNA
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(8883)
```

; OTHER INFORMATION: 'n' can be any nucleotide 'a', 'c', 'g' or 't'.
US-09-991-630-1

Alignment Scores:		6.73e-19	Length:	8883
Pred. No.:	Score:	416.50	Matches:	354
Percent Similarity:		30.73%	Conservative:	216
Best Local Similarity:		19.08%	Mismatches:	552
Query Match:		6.04%	Indels:	733
DB:		10	Gaps:	83
US-09-818-990B-2 (1-1320) x US-09-991-630-1 (1-8883)				
QY	2	GlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu	18	
DB	1460	CAGGAGCAATGGATCGCTCCACCTCACCTCAAGATTTCATAGAACCTTTGGGTCC	1519	
QY	19	SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla	35	
DB	1520	TTGTCTTTTGACATGACANANNTNTCTGGAATTAAGCCGACATGCTCTAGTATCCAA	1579	
QY	36	GluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly	55	
DB	1580	AAGCCATCAAGGACA-----TCACCAACTGCATTCACTGAA-----	1615	
QY	56	GlyGlnAspLeuProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGluSer	75	
DB	1616	-----GAAATCAGTACATCATGCTTAATGCGTCATT-----TCC	1651	
QY	76	ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln	95	
DB	1652	ACAAATCTGTG---TGCAGTGTAGATTATAATCACATCCAGCAGGTGGCACTCTG	1708	
QY	96	AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro	115	
DB	1709	GCTTTATACAGTGACTCTCTCTGATAGTACAGAAAGAGCCCGCAGCTTACC-----	1759	
QY	116	AsnPheCysGlnAspAsnProArgSerProThrSerLysGluSerProGlnGluAla	135	
DB	1760	-----GAGACTCCTTCACCTGTCTCTAGATATAACAGGTGCT	1798	
QY	136	LysArgProGlnTyrCysSerGluThr-----	144	
DB	1799	CTTAGGCTGAAGACATTTTACCAGCATAGAGGCTGATGTCAGAGCAGACCCCTTTTGG	1858	
QY	145	--GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluLeu	161	
DB	1859	TTCCAACAAGAAAAATTTGCTTCGAGCTGAACAGAACTGCCACCCACCTTAGCACATTA	1918	
QY	162	SerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsnHis	181	
DB	1919	CAGATCCAGTTTTCACATGATGCTCAATCCCTTTA---CCAAGGCG---GAGATGAGA	1972	
QY	182	LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp	201	
DB	1973	CGCGAGAGACTCAATGGACCATGATCCTGTATGATGAACAAT-----	2014	
QY	202	LeuSerGluArgArgGluArgSer-----	212	
DB	2015	-----CCCAAACTGGAACGACGTCTCTGTTGGCGGCACATTATGCCCTGAGCTGCCA	2068	
QY	213	-----IlePro	214	
DB	2069	GGCAAGCGGACCTTCACCTCAGTGGATGGCTTCTAGCTGATGGAGTAAGTGAAGA	2128	
QY	215	IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln	230	
DB	2129	GCCCTTACGTTAGCGAGGTGGCGGAATCCTAATAGACAAAAATGGGAAGTTGGAAGTG	2188	
QY	231	GlnGluAlaLysArgArgGluAla-----GluGlnAlaAlaSerGluAlaAlaGly	247	
DB	2189	CAGATGGCTGACAGCTTTGATGTCAGGCTTTTACCACCTGCATGAACCAATGATGCAGAT	2248	

QY	248	GlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGlu-----Glu	263	
DB	2249	GCGGATGTT-----CTCACATACAGGATAACTGGGTAGAG	2284	
QY	264	ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly	283	
DB	2285	CCCTATGGAGAAAGCACACATGACAGTGGAGTC-----CAGCACACAGTGGTTACGGGT	2338	
QY	284	ThrArgValGlnLeuAspCysIleValValGlyIleProProGlnValArgTyrTyr	303	
DB	2339	GAGACGCTCGACCTTCATGCCCTTCCACGGGTGTTCCAGATGCTTATAGCTGAT	2398	
QY	304	CysGluGlyLysGluLeuGluAsnSerPro-----AspIleHisIleValGlnAlaGly	321	
DB	2399	CTTCCAGGGAACACTGTTCTCTCAGCCATCAAGACAGACAGGCAATCTTAAACAATGGG	2458	
QY	322	AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys	341	
DB	2459	ACCTTAAGAATATTACAGGTTACG-----CCAAAGATCAAGTCTATTACCAATGT	2509	
QY	342	PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIle-----	358	
DB	2510	GTGCTGCCAACCCATCAGGGCGGACTTTCCAGTTTAAAGTTTCAGTTCAAAAGAAA	2569	
QY	359	--GluGlyValSerSerAspSerGlu-----GlyAspProAsn	371	
DB	2570	GGCCAAAGGATGTTGACCATGACAGGGAGGCGAGTGGATCTGCAGTTGGAGAACCCAC	2629	
QY	372	LysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerProThrThrSer	391	
DB	2630	TCCAGTGTTCCTTAAAGCAGCCAGCATCTTTGAAACTCTCTGCATCAGCTTTGACAGGG	2689	
QY	392	AlaValIleProProAlaValPro-----	399	
DB	2690	TCAGAGCTGGAACAAAGTCTCCGGTGTACATAGGAAGAACAAACATAGAGCTTAATA	2749	
QY	399	-----	399	
DB	2750	CATCGCGCGGTGGGATTCACCGCTCCGCGGATTCAGGGAGCATAGGAGCGACCTCCCT	2809	
QY	400	--GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGln---	417	
DB	2810	CTCTCTCTCGGAGATTGACCCGCAACGCTGGCGAGCATCTCTAGAAAAAGCCAAAAAG	2869	
QY	418	--SerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle-----IleAlaAlaPro	435	
DB	2870	ANTTCTGTGCCAAAAAGCAAGAAATACCACAGTAAGCCGAGTCCACACTGGCTGTTCCC	2929	
QY	436	ValPheThrLysMetLeuGlnAsnLeuSerAlaSer-----GluGlyGln	450	
DB	2930	CTCGTGGAACTCACTGACGAGGAAAGAGATCCCTCTGCATGATTCCTCCAGATGAAGAA	2989	
QY	451	LeuValValPheGluCysArgValLysGlyAlaPro-----SerProLysValGlu---	467	
DB	2990	TTTATGTTTCTGAAACAACTAAGGCTTCTGTCGCCAGGAAGTCCACCAACTGCTGACTCT	3049	
QY	468	-----TrpTyrArgGluGlyThrLeuIle-----	475	
DB	3050	GGACCACTAAATCATGGTTTTTATGACGAGTATAGCTTCTGCGCACAGAGTCTCAACTGTG	3109	
QY	476	-----GluAspSerProAspPheArg-----	482	
DB	3110	AATCCACAACACATACAAATCTGACACCTTCTCTGATTTCAAAATTTATTAGTGTAAACAAC	3169	
QY	483	-----IleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThr	499	
DB	3170	GGTACAGCTGTGACAAAAGAGATGAACCCATCCATAGCAAGAAAAATAGAAAGATACAACC	3229	
QY	500	-----LeuValIleAlaGluValPheAlaGlu-----AspSer-----	510	
DB	3230	AACCAAAACCAATCATTTATCTTCCATCAGTAGTGAATTCGAGATTCTGCTCAGGCA	3289	
QY	510	-----	510	

[illegible]

Db	4328	ACA	ACTCTCTCTTGACAAGTCCGCCACACAGCTCTGTCTCTACAAACAAATGGCTGCCACTCAGAAC	438
Qy	707	-----	***LysGlnValLysAlaProSerSerGlnThrPhe	718
Db	4388	AAG	GCACTGAAGTAGTATCAGGTGCCAGAAGTCTCTCAGCAGGGAAGAAGCAGCCCTTC	4447
Qy	719	SerLeuAlaArgProLysTyrPhePheProSerThr	-----AsnThrThr	733
Db	4448	ACC	AACCTCCTCTCCA-----GTGCTTCCTGACCAACATAAAGCAAGAGATCTAATACATTA	4501
Qy	734	-----AlaAlaThrValAlaProSer	---SerSerProValPheThrLeuSerSerThr	750
Db	4502	AAC	TTCTTGTCACGGGAACCCCCAGAGTCACAAAGTCCCTACTGCTACTGTCATCTGTCAATT	4561
Qy	751	ProGlnThrIleGlnArgThrValSerLysGluSer	-----	762
Db	4562	ATG	CTGAAACCCCAACGAACAAGATCCAAAGAAGCAAAAGACCAATAAAGGGCCCTCGG	4621
Qy	762	-----	-----	762
Db	4622	AAGA	CAACAAACGCAACACACCCCGCAGAGTTTCTGGCTATAGTGCATCTCA	4681
Qy	763	-----LeuLeuValSerHisProSerValGln	-----	771
Db	4682	GCT	CTAAACAACAGCTGATACCCCTTTGGCTTTTCAGTCAATCCCCACGACAGATGATGGT	4741
Qy	771	-----	-----	771
Db	4742	GGA	ATGTAAGTCGAGTTGTTATCACTCAACAACCTCTCTCTCGGCCATAACTGAAC TG	4801
Qy	772	-----	-----Thr	772
Db	4802	TTT	GAGAAGTACACCCAGACTTTGGAAATACAACAGCTTTTGGAAACAACGTTGTTGAGC	4861
Qy	773	LysSerProGlyGlyLeuSerIleGlnAsn	-----GluProLeu	785
Db	4862	AAAT	CACAGGAGAGTACCACAGTGAAGAAGCGCTCAGACACACCACCACTCCTCAGC	4921
Qy	786	-----ProProGlyProThrGluProThrProProPheThrPheSerIlePro	802	
Db	4922	AGT	GGGCGCCCCAGTGCCCACT---CCTTCCCAACCTCTCTTTACTAAGGGTGCGTT	4978
Qy	803	SerGlyAsnGln	-----PheGlnProArgCysValSerProIleProValSer	818
Db	4979	ACAG	ACGAAAGTCACATCAGCTTTCCAGATGACGTCAAATAGAGTGGTCACCATATAT	5038
Qy	819	ProThrSerArg	-----IleGlnAsnProValAla	828
Db	5039	GAAT	CTTCAAGGCACATACAGATCTGCAGCAACCTCAGCAGAGGCTAGCCCCAATCCT	5098
Qy	829	-----	-----PheLeuSerSerValLeuPro	835
Db	5099	GAG	ATCACTGAACCACTGACTCTCCCTCTAATCTGTCTTCCATCCACTCTGTGCGCA	5158
Qy	836	SerLeuProAlaIleProThrAsnAlaMet	***LeuProArgSerAlaProSerMet	855
Db	5159	GCACT	AGGGTAGATAAACACAGCAATCTAATGGAGCCCC-----	5200
Qy	856	ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAsn	875	
Db	5201	-----	-----TCTCCTGCGCCA-----	5212
Qy	876	IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAsp	895	
Db	5212	-----	-----	5212
Qy	896	ValArgProAsnGlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGlu	915	
Db	5213	-----GNAACAATATACGTCACTACTACCTCGAAGACCAATTGAGAAGGGC	5260	
Qy	916	IleGluPheArgLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAsp	935	
Db	5261	AAAAGGCCAGCAGTAAGCATGTCTCCCTC	5293	

```
QY 936 GluileProThrGlyLysCysIleAla----- 944
:::|::|
Db 5294 AGCCTCCAGAGCCAGCACTCATGCTCACACTGGAATACAGAGCATGCAGAAAG 5353
QY 945 ProIlePheAspLysArg----- 950
:::|::|::|::|::|
Db 5354 AGTGTTTTTGATAAGAAACCTGGTCAAAACCCCAACTTCCAAACATCTGCCTTACGCTCTCT 5413
QY 951 -----LeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
|::|::|::|::|::|
Db 5414 CTACCTAAGACTCTATTGAAAGACCAAGAAATAATTGGAGGAAGGCTGCAAGCTTTACA 5473
QY 966 -----CysLysIleValGlyIleProValProLys 975
|::|::|::|::|::|
Db 5474 GTTCCAGCTAATTCAGACGTTTTTCTCTCTGTGAGGCTGTGGAGACCCACTGCCCATC 5533
QY 976 ValTyrTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
:::|::|::|::|::|
Db 5534 ATCCACTGGACACAGAGTTTCATCAGGANTTGAATATCCCAAGGGACACAGAAAGCCGG 5593
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
:::|::|::|::|::|
Db 5594 TTCCACGTGCTTCCCAATGGCACCC-----TTGTCCATCCAGAGGTCAGTATTCAGGAC 5647
QY 1014 AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033
|::|::|::|::|::|
Db 5648 CGTGGACAGTACCTGCTCTGCTGATTAATCCACTGGCGTAGACCATTTTCATGCTCTCT 5707
QY 1034 LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArg 1053
|::|::|::|::|::|
Db 5708 TTGCTCTGTGTTTTTACCGG----- 5728
QY 1054 GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPro 1073
:::|::|::|::|::|
Db 5729 -----GCAAGGATTTTGGACACACATGTCAG----- 5755
QY 1074 HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAsp 1093
|::|::|::|::|::|
Db 5756 -----GAGATCACAGTTTCATCTTGGAGTACTGTGGAACTAAG 5794
QY 1094 CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVal 1113
|::|::|::|::|::|
Db 5795 TGCAGAGTGAGGGTATGCGAGCGCTACGGTTCTCTGGATACTTGCACCAACCAACGGTG 5854
QY 1114 LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
:::|::|::|::|::|
Db 5855 GTCTCAGAACGCCCAAGGAAGGAGCAGAAAGGCTGTGGGTAAACACCTGATGGA-----ACA 5908
QY 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn 1150
|::|::|::|::|::|
Db 5909 TTGATCATCTAATCTGAGCTCTTATGATCGTGGTTTTCACAGTGTTGGGCCACGAAC 5968
QY 1151 LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys 1170
:::|::|::|::|::|
Db 5969 CCATCTGGCCAGGATTCACGTGTTGGTTAAGATAACAAGTCATACA-----GCT 6016
QY 1171 AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro 1188
|::|::|::|::|::|
Db 6017 CCCCTGTCTATTATAGCAAAAGAGGCAAGCCATCGTTGGGTT---TTAGGTGGAAGT 6073
QY 1189 ValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAsp 1208
|::|::|::|::|::|
Db 6074 TTGAACCTGCCCTGCACCTGCAAAAGGAACTCCCGAGCTAGTGTTCCTGGGTCCTTTAT 6133
QY 1209 AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp 1224
:::|::|::|::|::|
Db 6134 GATGGGCTGAACATAAACCATTTGACGTTGACTCAITTCAGATTTTCTTGATCCAAAT 6193
QY 1225 ThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyr 1244
|::|::|::|::|::|
Db 6194 GGAAC-----CTGTATATAAGAGCATCGCTCCTCAGTCAAGTGGGGCACTTAT 6241
```

```
QY 1245 ThrLeuSerAlaLysAsnGluAlaGly----- 1253
|::|::|::|::|
Db 6242 GAGTGCATTGCCACCAGCTCCTCAGGCTCAGAGAGGGTAGTGATTCTTACTGTGGAA 6301
QY 1253 ----- 1253
Db 6302 GAGGAGAGACAATCCCGAGATAGAAACTGCCTCTCAGAAATGGACTGAGTGAATTG 6361
QY 1254 -----IleValSerCysThrAlaArgLeuAsp----- 1262
:::|::|::|::|::|
Db 6362 GGTGAGAAATTACTACTGAAGTCTCAGCTACTGGGATCCAAAGCCTAGAATAATCTGG 6421
QY 1263 -----IleTyrAlaGlnTrpHisGlnIleProProMetSer 1276
|::|::|::|::|::|
Db 6422 AGGCTGCCATCCAAGCTGTCATCGACCATGG---CACAGATGGCAGCCGAATCCAC 6478
QY 1277 ValArgProSerGlySerArg---TyrGlySerLeuThrSerLys 1290
|::|::|::|::|::|
Db 6479 GTCACCCCAATGGATCCTTGGTGGTGGTCAAGTACGACGGAATAA 6523

RESULT 15
US-09-991-630-5
; Sequence 5, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patent version 3.0
; SEQ ID NO 5
; LENGTH: 8883
; TYPE: DNA
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(8916)
; OTHER INFORMATION: n can be any amino acid
US-09-991-630-5

Alignment Scores:
Pred. No.: 6,73e-19 Length: 8883
Score: 416.50 Matches: 354
Percent Similarity: 30.73% Conservative: 216
Best Local Similarity: 19.08% Mismatches: 552
Query Match: 6.04% Indels: 733
DB: 10 Gaps: 83

US-09-818-990b-2 (1-1320) x US-09-991-630-5 (1-8883)
QY 2 GlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu 18
|::|::|::|::|::|
Db 1460 CAGGAGACAATGATGCTGCCTCCACTCACCAAGATTTTCATAGAACCTTTGGCTCC 1519
QY 19 SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla 35
|::|::|::|::|::|
Db 1520 TTGCTCTTGAACATGACANANNTTCTGGAATAAAGCCGACATGGTCTGTAGTATCCAA 1579
QY 36 GluProSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 55
:::|::|::|::|::|
Db 1580 AACCCATCAAGGACA-----TCACCAACTGCATTCACGTAA----- 1615
QY 56 GlyGlnAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSer 75
:::|::|::|::|::|
```


Db	1616	---GAAANTGACTACATCATGCTAAATGGCTCATTT-----TCC	1651
Qy	76	ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln	95
Db	1652	ACAAATCTCTGTG--TGCAGTGTAGATTATATATCATCATCCAGCCAGTGTGCCACTTCTG	1708
Qy	96	AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLysSerPheGluPro	115
Db	1709	GCTTTATACAGTCACTCTCTCTGTACTATAGAAAGGAGCCCGCAGCTTACC-----	1759
Qy	116	AsnPheCysGlnAspAsnProArgSerProThrSerSerLysGluSerProGlnGluAla	135
Db	1760	-----GAGACTCTCTTCATCTCTTCTAGATATAAAACAGGTGGCT	1798
Qy	136	LysArgProGlnTyrCysSerGluThr-----	144
Db	1799	CTTAGGCGCTGAACACATTTTACCAGCATAGAGGCTGATGTCAGACAGCACCCTTTTGG	1858
Qy	145	---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluGluLeu	1616
Db	1859	TTCCAACAAGAAAAAATTGCTTGCAGCTGAACAACAAGCTGCCACCACACTTAGCACATTA	1918
Qy	162	SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis	181
Db	1919	CAGATCCAGTTTTTCCACTGATGCTCAAAATCGGTTTTA---CCAAAGGGCG--GAGATGAGA	1972
Qy	182	LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSerPheSerAsp	201
Db	1973	GCGGAGAGACTCAAAATGGACCATGATCTCTGATGATGACAAT-----	2014
Qy	202	LeuSerGluArgArgGluArgSer-----SerValPro	212
Db	2015	---CCCAAACTGGACGCACACTCTCTGGTGGCGGCACACTATTGCCCTGAGCTGTCCA	2068
Qy	213	-----IlePro	214
Db	2069	GGCAAAAGGGGACCCCTTCACTCACTTGGAAATGGCTTCTAGCTGATGGGAGTAAAGTGAGA	2128
Qy	215	IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln	230
Db	2129	GCCCCCTTAGTTCGCGAGGATGGCGAAATCCTTAATAGACAAAAATGGGAAGTTGGAAGT	2188
Qy	231	GlnGluAlaLysArgArgGluAla-----GluGlnAlaAlaSerGluAlaAlaGly	247
Db	2189	CAGATGGCTGACAGCTTTCATGCAGCTTCTTACCACCTGCATAAGCACCAATGATGCAGAT	2248
Qy	248	GlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGlu-----Glu	263
Db	2249	GCGGATGTT-----CTCACATACAGGATAAAGTGTGGTAGAG	2284
Qy	264	ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly	283
Db	2285	CCCTATGGAGAAGACACATGACAGTGGAGTGC-----CAGCACACAGTGGTTACGGGT	2338
Qy	284	ThrArgValGlnLeuAspCysIleValValGlyIleProProProGlnValArgTyrTyr	303
Db	2339	GAGACGCTCGACCTTCCATGCCTTTCCACGGGTGTCAGATGCTCTTATTAGCTGGATT	2398
Qy	304	CysGluGlyLysGlnLeuGluAsnSerPro-----AspIleHisIleValGlnAlaGly	321
Db	2399	CTTCCAGGGAACACTGTGTCTCTCAGGCCATCAAGAGACAGGCAAAATCTTAAACAATGGG	2458
Qy	322	AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys	341
Db	2459	ACCTTAGAATATTTACAGGTTTACG-----CCAAAAGATCAAGGTCTATTACCAATGT	2509
Qy	342	PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIle-----	358
Db	2510	GTGGCTGCCAACCCATCAGGGGCCACACTTTTCCAGTTTAAAGTTTCAGTTCAAAAGAAA	2569
Qy	359	---GluGlyValSerSerSerAspSerGlu-----GlyAspProAsn	371
Db	2570	GGCCAAAGAGTGTGTTGACATGCACAGGAGGCGCAGGTGGATCTGGACTTGGAGAACCCAC	2629

QY 568 -----ProSerValGluGlnPro----- 573
Db 3710 GCTAACAAAAATGTGAGCCAAGTTCCAGCCACAGAGTACCCTGGGTGTCACACACATGT 3769
QY 573 ----- 573
Db 3770 CCTTCGCAGAGGGGCTCACAGTGGCTACTGCGACACTGTGTCAGTTCCAAGTTTCATCCACC 3829
QY 574 -----ProLysProLysLeuGluGlyVal----- 583
Db 3830 AGTGCCCTCCCAAACTAATAATGTGGGTCTATAGCAGAAAGTCTACCAAGAGTCTACCACTGGTGC 3889
QY 583 ----- 583
Db 3890 AAGAAACCACTGTTACTATTAAAGCAACAACAAATGTAGATATTGAGATAATAACAACC 3949
QY 584 -----AsnHisAsnGluProProArgSerSerSerArgIle 594
Db 3950 ACTACAAAATATTCGGGAGGGAAAGTAACCGTGTATCTCTAGCGAAGCAAGCATGACT 4009
QY 595 GlyLeuArgValHisPheAsnLeu-----ProGluAspAspLysGlySerGluAla 611
Db 4010 TCTGCTCCACACTGTATCCCTGGGGAATCTCTCTAGACAATAGTGGT----- 4060
QY 612 SerSerGluAlaGlyValValThrArgGlnThrArgProAspSer***GlnGluArg 631
Db 4061 -----CACCTGACCATGCGTGGGACCATCCAACTGGGAAGATTCAGTGGAA----- 4108
QY 632 PheAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProPro--- 650
Db 4109 -----ACAACACCACTTCCAGC-----CCCTCAGCACACCCCTCAATA 4147
QY 651 ProValLeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeu-----HisAsn 668
Db 4148 CCAACAGCACAAAATCTCAAGAGAGAAACTCCCTGTGCACCATCTTTGTAAATTAAC 4207
QY 669 Gln-----ValLeuLeuGluGlnHisGlnLeu-----GlnAsnProPro 681
Db 4208 CAGAAGAAGAGGGGATGTTAAAGAATCCATATCAATTCGGTTTACAAAAGAACCCAGCC 4267
QY 682 ProSerSerProLys-----GluPheProPhe***Met 692
Db 4268 GCAAGGCTTCCCAAAATAGCTCTCTTTACCCACAGGTGAGAGTTCCTCCCTCAGATCTCT 4327
QY 693 ThrValLeuAsnSerAsnAlaProProAlaValThrThrSer----- 706
Db 4328 ACAACTCTTGACAAGTCGCGCCACAGCTGTGCTACACAAATGGCTGCCACTCAGAAC 4387
QY 707 -----***LysGlnValLysAlaProSerSerGlnThrPhe 718
Db 4388 AAGGCACCTGAAGTAGTATCAGGTGCGCAGAAGTCTCTCAGCAGGGAAGACGCCCTTC 4447
QY 719 SerLeuAlaArgProLysTyrPhePheProSerThr-----AsnThrThr 733
Db 4448 ACCAACTCCTCTCCA-----GTGCTTCTTAGCACCATAAGCAAGAGATCTAATACATTA 4501
QY 734 -----AlaAlaThrValAlaProSer---SerSerProValPheThrLeuSerSerThr 750
Db 4502 AACTCTTGTCACCGGAAACCCACACAGTGACAAAGTCTACTGTCTACTGTCATCTGTCAAT 4561
QY 751 ProGlnThrIleGlnArgThrValSerLysGluSer----- 762
Db 4562 ATGCTGAAACCCCAACGAACAAGATCCAAAGAACCAACAAATAAAGGGCCCTCGG 4621
QY 762 ----- 762
Db 4622 AAGAACAGAAACACGCAAAACACACCCAGCGAGGTTTCTGTGCTATAGTGCATCTCA 4681
QY 763 -----LeuLeuValSerHisProSerValGln----- 771
Db 4682 GCTCTAACACAGCTGTATACCCCTTGGCTTTGAGTCATCTCCCCACGACAAAGATGATGGT 4741
QY 771 ----- 771

Db 4742 GGAATGTAAAGTGCAGTGTCTTATCTACTCAACAACCTCTCTCTGGCCATAACTGAAC TG 4801
QY 772 -----Thr 772
Db 4802 TTTGAGAAGTACACCCAGACTTTTGGGAAATACAAACAGCTTTGGAAACAACAGTGTGTTGAGC 4861
QY 773 LysSerProGlyGlyLeuSerIleGlnAsn-----GluProLeu----- 785
Db 4862 AAATCACAGAGAGTACCACAGTGAAGAGCCCTCACACACACACACTCTCTCAGC 4921
QY 786 -----ProProGlyProThrGluProThrProProPheThrPheSerIlePro 802
Db 4922 AGTGGGCGCCCGCCAGTGCCTACT---CCTTCCCACTCTCTTTACTAAGGTTGTGTT 4978
QY 803 SerGlyAsnGln-----PheGlnProArgCysValSerProIleProValSer 818
Db 4979 ACAGACAGCAAAAGTTCACATCAGCTTTCCAGATGACGTCAAAATAGAGTGGTCACCATATAT 5038
QY 819 ProThrSerArg-----IleGlnAsnProValAla----- 828
Db 5039 GAATCTTCAAGGCACAATACAGATCTCGCAACCCCTCAGCAGAGGCTAGCCCCAATCCT 5098
QY 829 -----PheLeuSerSerValLeuPro 835
Db 5099 GAGATCATAACTGGAACCACTGACTCTCCCTCTAATCTGTTTCCATCCACTCTCTGTGCCA 5158
QY 836 SerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMet 855
Db 5159 GCATAGGGGTAGATAAACCAACACAGAAATCTAAATGGAAGCCC----- 5200
QY 856 ProSerGlnGlyLeuAlaLysAsnThrLysSerProGlnProValAsnAspAsn 875
Db 5201 -----TCTCCCTGGCCA----- 5212
QY 876 IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysIleThrPheSerAsp 895
Db 5212 ----- 5212
QY 896 ValArgProAsnGlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGlu 915
Db 5213 -----GAACACAAATATCAGCTCAAGTCATACTCCGAAACCATTTGAGAAGGCG 5260
QY 916 IleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAsp 935
Db 5261 AAAAGGCCAGCATGTAACATGTCCTCC-----CACCTC 5293
QY 936 GluIleProThrGlyLysCysIleAla----- 944
Db 5294 AGCCTTCCAGAGGCCAGCACTCATGCTCCTCAGCTGGAATACACAGAACGATGCAAAAAG 5353
QY 945 ProIlePheAspLysArg----- 950
Db 5354 AGTGTGTTTGTATAAGAAACCTGGTCAAAACCCAACTTCCAAACATCTGCTCTAGTCTCT 5413
QY 951 -----LeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
Db 5414 CTACCTAAGACTCTATTGAAAACCCCAAGAAATAATTGGAGGAAGGCTGCAAGCTTTTACA 5473
QY 966 -----CysLysIleValGlyIleProValProLys 975
Db 5474 GTTCCAGCTAATTCAGAGCTTTTCTCTTGTGAGGCTGTGGAGACCCCACTGCCCATC 5533
QY 976 ValTyrTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
Db 5534 ATCCACTGGACAGAGTTTCATCAGGANTTGAATATATCCCAAGGACACACAAAAGCGG 5593
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
Db 5594 TTCACAGTGTTCCTCAATGGCACCC-----TTGTCCATCCAGAGGGTCAATTCAGGAC 5647
QY 1014 AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033

```
Db 5648 CGTGACAGTACCTGCTGCTGCATTTAATCCACTGGCGGTAGACCAATTTTCATGCTCTCT 5707
Qy 1034 LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArg 1053
Db 5708 TTGCTGTGGTGTTTTACCCG----- 5728
Qy 1054 GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArgPro 1073
Db 5729 -----GCAAGGATTTTGACAGACATGTC AAG----- 5755
Qy 1074 HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAsp 1093
Db 5756 -----GATACACAGTTCATTTTGAAGTACTGCTGGAACATAAG 5794
Qy 1094 CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVal 1113
Db 5795 TGCAGGTGGAGGTATGCCAGGCTACGGTTCTCGTACACTTGC AACCACACGGTG 5854
Qy 1114 LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
Db 5855 GTCTCAGAAACGGCCAGGGAAGCAGAAAGGTCTGGGTAACACCTGATGGA-----ACA 5908
Qy 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn 1150
Db 5909 TTGATCATCTATAAFTCAGTCTTTATGATCGTGGTTTACAAAGTGTGCCAGCAAC 5968
Qy 1151 LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValAlaLysGluValLysLys 1170
Db 5969 CCATCTGCCAGGATTCATCTGTTGTTAAGATACAGTCATCACA-----CCT 6016
Qy 1171 AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro 1188
Db 6017 CCCCTGTCTATTATAGACAAAGAGGCAAGCCATCGTTGGGGTT---TTAGTGGAAGT 6073
Qy 1189 ValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAsp 1208
Db 6074 TTGAAACTGCCCTGCACCTGC AAAAAAGGAAGTCTCCAGCCCTAGTGTTCCTGGTCTCTTAT 6133
Qy 1209 AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp 1224
Db 6134 GATGGACTGAACCTAAACCATTCGACTGATCCAGATTTTCTGTATCCAAAT 6193
Qy 1225 ThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyr 1244
Db 6194 GGAAC-----CTGTATATAAGAAGCATCGCTCCTTCAGTGAGGGGCACCTTAT 6241
Qy 1245 ThrLeuSerAlaLysAsnGluAlaGly----- 1253
Db 6242 GAGTGCAATTGCCACAGCTCCTCAGGCTCAGAGAGAGGGTAGTGATTTACTGTGGAA 6301
Qy 1253 ----- 1253
Db 6302 GAGGAGAGACAATCCCGAGGATAGAAACTGCCCTCAGAAATGGACTGAGGTGAATTTG 6361
Qy 1254 -----IleValSerCysThrAlaArgLeuAsp----- 1262
Db 6362 GGTGAGAAATTACTACTCACTGCTCAGCTACTGGGGATCCAAAGCCTAGAAATAATCTGG 6421
Qy 1263 -----IleTyrAlaGlnTrpHisHisGlnIleProProPrometSer 1276
Db 6422 AGGTGCCATCCAAAGGCTGTCTACACCATGG---CACAGAAATGGCGCGCAATCCAC 6478
Qy 1277 ValArgProSerGlySerArg---TyrGlySerLeuThrSerLys 1290
Db 6479 GTCTACCCCAATGATCCTTGGTGGTGGTACGTGACCGAATAA 6523
```

Search completed: November 30, 2002, 22:52:06
Job time : 319 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:09:50 ; Search time 2652 Seconds
(without alignments)
8061.102 Million cell updates/sec

Title: US-09-818-990B-2
Perfect score: 6890
Sequence: 1 MODSIEASTSISQLRESY.....MESTMVYSCSRVSDEL 1320

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09818990/runat_26112002_093405_23231/app_query.fasta_1.1479
-DB=EST -QFWT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGL1.1.1996 @runat_26112002_093405_23231 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hctc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hctc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1375.5	20.0	3865	11	AF077041	Homo sapi
2	1298.5	18.8	1285	11	AK010350	Mus muscu
3	965	14.0	614	14	BM968035	LM24HM007
4	910.5	13.2	901	14	BQ881261	AGENCOURT
5	908	13.2	595	13	BJ071900	BJ071900
6	887	12.9	943	13	BI691193	603314585
7	877	12.7	868	12	BG827059	602749109
8	869.5	12.6	921	13	BI758769	60323796
9	858.5	12.5	717	14	BQ206762	UI-R-D21-
10	841	12.2	553	12	BF126187	601650450
11	831	12.1	828	13	BI526367	602925663
12	799	11.6	612	14	BM795050	K-EST0076
13	779	11.3	481	9	AA179599	zP49f09.f
14	768	11.1	654	10	BB548628	BB548628
15	758	11.0	772	13	BG923162	602824001
16	733	10.6	579	13	BJ041619	BJ041619
17	728	10.6	674	13	BI414869	602990942
18	711	10.3	1050	17	CNS05HM4	AL337765 Tetraodon
19	704	10.2	2209	11	AK017447	AK017447 Mus muscu
20	680	9.9	1101	17	CNS05AZS	AL329185 Tetraodon
21	679.5	9.9	874	12	BG423154	602450365
22	678.5	9.8	563	13	BJ069302	BJ069302
23	671.5	9.7	413	14	BQ559749	BQ559749 H4060802-
24	663	9.6	804	13	BI662601	BI662601 603302549
25	662	9.6	506	9	AL709148	DRFZP686L
26	658	9.6	376	10	AW918754	AW918754 EST350058
27	652	9.5	503	17	AQ621045	HS_2221.B
28	646	9.4	444	13	BJ070056	BJ070056
29	643	9.3	594	13	BJ065434	BJ065434
30	639.5	9.3	969	14	BQ722262	AGENCOURT
31	638	9.3	804	14	BQ202926	BQ202926 UI-R-D01-
32	633	9.2	438	10	AW315222	AW315222 12352 MAR
33	631.5	9.2	706	9	AI745576	WC34901.x
34	621.5	9.0	713	9	AA887587	AA887587 OJ54H02.S
35	619	9.0	483	12	BE855373	BE855373 ux51h11.y
36	614.5	8.9	681	10	BB045059	BB045059 BB045059
37	598	8.7	616	12	BF238982	BF238982 601905239
38	597.5	8.7	558	9	AA128252	2130b11.f
39	580.5	8.4	670	10	BB486332	BB486332
40	575	8.3	531	9	AA446042	zw66b03.f
41	572.5	8.3	707	9	AL598867	DRFZP3130
42	572	8.3	501	10	BE448779	BE448779 ut83q05.y
43	571.5	8.3	569	9	AA187765	AA187765 zp61a10.f
44	569	8.3	513	10	BB134094	BB134094
45	562	8.2	506	9	AA285584	vb83f11.f

ALIGNMENTS

RESULT 1
AF077041
LOCUS
DEFINITION Homo sapiens
ACCESSION AF077041
VERSION AF077041.1
KEYWORDS HTc.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3865)
REFERENCE Liu,T., Zhang,J., Ye,M., Zhang,Q., Fu,G., Zhou,J., Wu,J., Shen,Y.,
AUTHORS Yu,M., Chen,S., Mao,M. and Chen,Z.
TITLE Human SIH002 gene

AF077041 3865 bp mRNA linear HTC 22-MAY-2001
Homo sapiens SIH002 mRNA, complete cds.
AF077041.1 GI:4689129
HTc.
Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3865)
REFERENCE Liu,T., Zhang,J., Ye,M., Zhang,Q., Fu,G., Zhou,J., Wu,J., Shen,Y.,
AUTHORS Yu,M., Chen,S., Mao,M. and Chen,Z.
TITLE Human SIH002 gene

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3865)
AUTHORS Liu,T.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-1998) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P.R. China
FEATURES Location/Qualifiers
source 1..3865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NB4"
648..1805
/codon_start=1
/product="SIH002"
/protein_id="A0D2777A.1"
/db_xref="GI:4689130"
/translation="MAPFFEMKLHKYIFGMPVTFCTRVAGNPKIYFKDGKQIS
PKSDHYTIORLDGTCSLHTASTLDDNGYITMAANPOGRI SCTGRIMQAVNQGR
SPRSPGHQHVRRPSRSDSDENEP IQERFFRPHFLOAPGDLTVQEGKLCRMCKV
SGLPTDLSQLDGKVPKRPDSAHKMLVRENGVHSLIIEPTVSRDAGLYTCTIATNRAG
NPSLELVAAKKAHPPVPIELKONTGVADGYPVLECRVLGVPPPIFWKKENESL
THSDRSMQDHNHGIICLIQATGEDAGWYIVSAKNEAGIVSCVTRALDVIYQWQQ
SQSTKPKKVPKPSASRAALSQDGLDKAAFPQEPANPSHLTLNLTALVESEDL"
BASE COUNT 1238 a 739 c 788 g 1100 t
ORIGIN
Alignment Scores:
Pred. No.: 9,19e-113 Length: 3865
Score: 1375.50 Matches: 260
Percent Similarity: 75.93% Conservativeness: 68
Best Local Similarity: 60.19% Mismatches: 91
Query Match: 19.96% Indels: 13
DB: 11 Gaps: 4
US-09-818-990b-2 (1-1320) x AF077041 (1-3865)
Qy 901 GlnGluTyrLysIleSerSerPheGlnGlnArgLeuMetAsnGluIleGluPheArgLeu 920
Db 514 CAGGAATACAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAGATCAGGCTA 573
Qy 921 GluArgThrProValAspGluSerAspGluIleGlnHis-----AspGluIlePro 938
Db 574 GAAAGGTCTCCTATGGATGAATCATGTGTGATGAACGTTTCACGTACGCGCAGATCGCT 633
Qy 939 ThrGlyLysCys--IleAlaProIlePheAspLysArgLeuLysHisPheArgValThrG 958
Db 634 CGTGGAAACTGGAAATGGCACCATTCTTTGAGATGAAGCTGAAACATTACAAAGATCTTG 693
Qy 958 luGlySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrT 978
Db 694 AGGAATGCCAGTAACCTTCACATGTAGATGGCTGGAAATCCAAAGCCAAAGATCATTT 753
Qy 978 rpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluG 998
Db 754 GGTGTTAAAGATGGGAAGCAGATCTCTCAAGAGCTGATCATACACCATTCAAAGAGATC 813
Qy 998 lyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrT 1018
Db 814 TCGATGGACCTGCTCCTCCATACACAGCTCCACCTCCTAGCTGATGGGAATTATA 873
Qy 1018 hrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln 1038
Db 874 CAATATGGCTGCAACACCTCAGGCGGCATCAGTTGTACTGGAGCGCTAATGTACAGG 933
Qy 1038 erLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis----- 1052
Db 934 CTGTCAACCAAGAGGTGGAAGTCCCGGCTCCCTCAGGCGCATCCTCATGTCAAGAGGC 993
Qy 1053 --ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheA 1072
Db 994 CTCGTTCTAGATCAAGGACAGCTGAGACGCAAAATGAACCAATTTCAGGAGCGGATCTTCA 1053

Qy 1072 rgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgL 1092
Db 1054 GACCTCACTTCTTGCAGGCTCTCGAGATCTGACTGTTCAAGAGGAAACTCTGCAGAA 1113
Qy 1092 euAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnP 1112
Db 1114 TGGACTGCAAAAGTCAGTGGGTACCAACCCAGATCTAAGCTGGCAACTAGTAGAAAGC 1173
Qy 1112 roValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1132
Db 1174 CCGTACGCCCTGACAGTCTCACAAGATGCTGGTGGCGTGAAGACGGGTGCACCTCTCTGA 1233
Qy 1132 euIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysT 1152
Db 1234 TCATAGAGCCAGTCAGCTCAGCTGATGCCGCATCTACACATGTATAGCTACCAACCGAG 1293
Qy 1152 hrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaP 1172
Db 1294 CAGGACAGAACTCATTCAGCCTGGAGCTTGTGTGCTGTCTAAAGAACACACAAACCC 1353
Qy 1172 roValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeu 1192
Db 1354 CTGTGTTATTGGAAGCTCCAAAACACAGAGTTGCTGATGGGTACCCAGTGGCGGTGG 1413
Qy 1192 luCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrI 1212
Db 1414 AATGCTGTATTGGGAGTGCACACCTCAGATATTTTGGAAAGAAATAATGAATCAC 1473
Qy 1212 leProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeu 1232
Db 1474 TCACCTCAGACGACTGACCGAGTGAAGTGCACAGGACAAACCGGCTTACATCTGCCTGC 1533
Qy 1232 euIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluA 1252
Db 1534 TCATTGAGGGAGCCCAAAAGAGTGTGGTGGTATCTGTGTGACGCCAAGAATGAAG 1593
Qy 1252 laGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis-----G 1270
Db 1594 CAGGGATTGTCTCTGTACTGCCAGGCTGGAGCTTTACACCCAGTGGCAGTGCAGAGTCAC 1653
Qy 1270 InIleProProMetSerValArgProSerGlySerArgTyrGlySerLeuThrSerL 1290
Db 1654 AGACACCAAGCAACCAAAAGTACGGCCCTCAGCAGCTGCGTATGACGACACTTTTGGACC 1713
Qy 1290 ysGlyLeuAspIlePheSerAlaPheSerSerMetGluSerThrMetValTyrSerCys 1310
Db 1714 AGGACTAGACATCAAGCAGCGTTCCAACCT---GAGGCCAACCCATCTCACCTGACAC 1770
Qy 1310 erSerArgSerValValGluSerAspGluLeu 1320
Db 1771 TGAATCTGCTTGGTAGAAGTGAAGGACCTG 1802
RESULT 2
AK010350 1285 bp mRNA linear HTC 19-JAN-2002
LOCUS Mus musculus ES cells cDNA, RIKEN full-length enriched library,
DEFINITION clone:2410003B16:homolog to CGI-151 PROTEIN, full insert sequence.
ACCESSION AK010350.1 GI:12845726
VERSION AK010350.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2410003B16.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2


```

Db 370 AGAAGGCGCTCGCTCCGATCAGCGGACAGTGGAGATGAAACAGAGCCATTTCAGGACGA 429
Qy 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeu 1089
Db 430 TTCTTCAGACTCACTTCCCTGCAGGCTTCCTGGAGACCTGACCGTTCAGGAGGCAAGCTC 489
Qy 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsn 1109
Db 490 TGCAGGATGACTGCAAGTCACTGGGTACCAACCCAGATCTCAGCTGGCACTAGAT 549
Qy 1110 GlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
Db 550 GGAAGGCCATAGACCCCGACAGCTGCTCACAAGATGCTGGTCCGTGAGAATGGGGTCCAC 609
Qy 1130 SerLeuLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThr 1149
Db 610 TCCCTCATTTAGAGCAGTCACTCCCGGACGCCGCGCATCTACACATGATTGGCCACC 669
Qy 1150 AsnLysThrGlyGlnAnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
Db 670 AACAGCAGCAGACAGACTCGTTTAACCTGGAGCTTGTGCTGCTAAGGAAGCACAC 729
Qy 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
Db 730 AAGGCCCTGTGTTTATGGAGAAGCTACAGAACACGGGGGTGCTGATGGATACCCAGTG 789
Qy 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTrpTrpLysLysAspAsn 1209
Db 790 CGGCTGGAATGCCGTGCTCGGAGTGGCCGCCCTCAGATATTTTGGAGAAGAAAAAT 849
Qy 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTrpAla 1229
Db 850 GAATCGCTCACTCACACCACTGACGAGTAGAGCATGACAGGATATCATGGCTATCATC 909
Qy 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTrpTrpLysLeuSerAlaLys 1249
Db 910 TGCTGTCTCATCCAGGAGGACCAAGAAAGACGCTGGGTGTACACTGTGTGCGGCAAG 969
Qy 1250 AsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTrpAlaGlnTrpHisHis 1269
Db 970 AACGAACGAGCATGTGTCTGTGACATGCCCCGGCTGGATGTCTACACCGATGGCACAG 1029
Qy 1270 -----GlnIleProProMetSerValArgProSerGlySerArgTrpGlySerLeu 1287
Db 1030 CAGCCACAGACCAACCAAGCAAAAAAGTAGACGGCCCTCGGCCAGTCTGCTACGCACTT 1089
Qy 1288 ThrSerLysGlyLeuAspIlePheSerAlaPheSerSerMetGluSerThrMetValTrp 1307
Db 1090 TCGACACGAGGATAGACATCAAGCCGCTTCCAACT--GAAGCCAGCCCATCTCAC 1146
Qy 1308 SerCysSerArgSerValGluSerValGluSerAspGluLeu 1320
Db 1147 CTGACTTTGAACACCGCTTGGTAGAAGTGAAGACCTG 1185

RESULT 3
BM968035
LOCUS
DEFINITION
  LM24HW00787 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone
  LM-24-HW-007-87 (5'), mRNA sequence.
ACCESSION
  BM968035
VERSION
  BM968035.1 GI:19562460
KEYWORDS
  EST.
SOURCE
  cow.
  Bos taurus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovine; Bos.
REFERENCE
  1 (bases 1 to 614)
  Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,
  Sun,S.S. and Cheong,I.C.
  Gene Expression Profiling of the Bovine skeletal muscle
  Unpublished (2002)

```

```

COMMENT
  Contact: Dr. Du-Hak Yoon
  National Livestock Research Institute, RDA
  564 Omoekchun-dong, Suwon, 441-350, Korea
  Tel: 82 31 290 1593
  Fax: 82 31 290 1792
  Email: dhyoon@rda.go.kr
  Insert Length: 614 Std Error: 0.00
  Seq primer: CAGGAACAGCTATGAC
  POLYA-No.
FEATURES
  source
  1..614
  /organism="Bos taurus"
  /db_xref="taxon:9913"
  /clone="LM-24-HW-007-87 (5')"
  /clone_lib="Bos taurus LM-24-HW cDNA library"
  /sex="six males mixed"
  /tissue_type="longissimus dorsi"
  /cell_type="myocyte"
  /dev_stage="24 months old"
  /lab_host="XLI-BlueMRF/strain"
  /note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site_1:
  EcoRI; Site_2: Xho I"
  BASE COUNT 150 a 188 c 159 g 117 t
  ORIGIN
  Alignment Scores:
  Pred. No.: 7,01e-77 Length: 614
  Score: 965.00 Matches: 190
  Percent Similarity: 93.66% Conservative: 2
  Best Local Similarity: 92.68% Mismatches: 10
  Query Match: 14.01% Indels: 4
  DB: 14 Gaps: 2
  US-09-818-990B-2 (1-1320) x BM968035 (1-614)
  Qy 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeu 1089
  Db 2 TTCTTCGACACACATTTCCTGCAGGCTCCTGGGATATGGTGGCTCAGGAGGCGGCTC 61
  Qy 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsn 1109
  Db 62 TGCGGCTGGACTGTAAAGTGTAGTGGCTTACCGCCCCGGAGCTGACGTGGCTGCTCAAT 121
  Qy 1110 GlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
  Db 122 GGCAGGCTGTCTCCCGACACCTCCCAAGATGCTGTGTCAGGAGACGGAGTCCAC 181
  Qy 1130 SerLeuLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThr 1149
  Db 182 TCTGTGCTCATTTGACCGCTCACTCAACGCGACGCGGGACCTACACCTGCATTGCTACC 241
  Qy 1150 AsnLysThrGlyGlnAnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
  Db 242 AACAAAACCTGGCGAGAATTCCTTTAGCTTGGAGCTACTAGTAGCCCAAGAAAGTAAG 301
  Qy 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
  Db 302 AAGSCACCTGTAATCTCGAGAAACTCAGAACAGCGGTGTTCAGAAAGGCCACCCCGTG 361
  Qy 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTrpTrpLysLysAspAsn 1209
  Db 362 AGACTGGAGTGGCGGGTGATAGGCATGCCCTCCCGCCCTGTCTACTGGAAGAAAGACAAC 421
  Qy 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTrpAla 1229
  Db 422 GAGACCATCCCTTTCACACAGAGAGAGATCAGATGCACGACGACACACTGGGTATGTC 481
  Qy 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTrpTrpLysLeuSerAlaLys 1249
  Db 482 TGCCTCTCTATTCAGCCAGCAAAAAATCAGATGCTGGATGGTACACACTGTACGCAAG 541
  Qy 1250 AsnGluAlaGlyIleValSerCysThr-AlaArgLeuAspIleTrpAlaGlnTrpHisHis 1269
  |||||||||||||||||||||||||||

```

```
Db 542 AC-GAAGCGGGCATCGTCTCGACTGGCA---GCTGATATATACGCTATG---GATCA 594
QY 1269 sGlnLeuPro 1273
|::: |||||
Db 595 CAAACCCGCCCA 607

RESULT 4
BQ881261
LOCUS
DEFINITION BQ881261 901 bp mRNA linear EST 16-AUG-2002
AGENCOURT.7978307 Lupski dorsal_root_ganglion Homo sapiens cDNA
clone IMAGE:6185000 5', mRNA sequence.
BQ881261
VERSION BQ881261.1 GI:22273269
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 901)
NIH-MGC http://imgc.ncbi.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13575 row: b column: 09
High quality sequence stop: 610.
FEATURES
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6185000"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-GACTAGTCTTATGATCGAGCGGCCGCT(15)-3' and
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 260 a 232 c 223 g 186 t
ORIGIN

Alignment Scores:
Pred. No.: 1,07e-71 Length: 901
Score: 910.50 Matches: 169
Percent Similarity: 79.39% Conservative: 39
Best Local Similarity: 64.50% Mismatches: 50
Query Match: 13.21% Indels: 4
DB: 14 Gaps: 2

US-09-818-990B-2 (1-1320) x BQ881261 (1-901)
QY 1062 AsPlyGluProGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAsp 1081
::: |||||:::
Db 34 GAAATGAACCAATTCAGGAGCATCTTCAGACCTCCTCTTCGAGGCTCTGGAGAT 93
::: |||||:::
QY 1082 MetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeuProPro 1101
::: |||||:::
Db 94 CTGACTGTTCAAGAAGGAAACTCTGCAGAAATGGACTGCAAGTCAAGTGGTTTACCAACC 153
```

```
QY 1102 ProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMet 1121
|::: |||||:::
Db 154 CCAGATCTAAGCTGGCAACTAGATGGAAGCCCGTACGCCCTCACAGTGTCTACAGATG 213
|::: |||||:::
QY 1122 LeuValArgGluThrGlyValHisSerLeuLeuLeuIleAspProLeuThrGlnArgAspAla 1141
|::: |||||:::
Db 214 CTGGTGGTGGAGAACGGGTGCACCTCTCTGATCATAGAGCAGCTCAGCTCAGCTGATGCC 273
|::: |||||:::
QY 1142 GlyThrTyrLysCysIleAlaThrAsnLysThrGlyClnAsnSerPheSerLeuGluLeu 1161
|::: |||||:::
Db 274 GGCATCTACACATGTATAGCTACCAACCGAGCAGAGAACTCATTCAGCCTGGAGCTT 333
|::: |||||:::
QY 1162 SerValValAlaLysGluValLysAlaProValIleLeuGluLysLeuGlnAsnCys 1181
|::: |||||:::
Db 334 GTGGTTGCTGCTAAAGAACACACAAACCCCTGTGTTTATGAGAAGCTCCAAACACA 393
|::: |||||:::
QY 1182 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProPro 1201
|::: |||||:::
Db 394 GGAGTTGCTGATGGGTACCCAGTGGCTGGAATGCTGTTATTTGGAGTGCACCACT 453
|::: |||||:::
QY 1202 ValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet 1221
|::: |||||:::
Db 454 CAGATATTTTGGAGAAAGAAATGAATCACTCACTCACAGCAGCTGACCGAGTGAGCATG 513
|::: |||||:::
QY 1222 HisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysSerAspAla 1241
|::: |||||:::
Db 514 CACGAGCAGCAACACGCTACATCTGCTCTCTATTCAGGAGCCACAAAGAGATGCT 573
|::: |||||:::
QY 1242 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeu 1261
|::: |||||:::
Db 574 GGCTGTATATCTGTCAGCCAAAGATGAAGCAGGAGTGTCTCTGCTACTGCCAGGCTG 633
|::: |||||:::
QY 1262 AspIleTyrAlaGlnTrpHisHis-----GlnIleProProProMetSerValArgPro 1279
|::: |||||:::
Db 634 GACGTTTACACCCAGTGGCATCGACGATCACAGCAGCACCAAGCAAGTACGCCCC 693
|::: |||||:::
QY 1280 SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSer 1299
|::: |||||:::
Db 694 TCAGGAGTGCCTATGCGAGCAGCTTCGGACCGAGGACTAGACATCAAGCAGCGTTCCCA 753
|::: |||||:::
QY 1300 SerMetGluSerThrMetValTyrSerCysSerSerArgSerVal-ValGluSerAspG1 1319
|::: |||||:::
Db 754 ACT---GAAGGCAACCCATCTCACTGACACTGAAATACTGCTTGGGTAGAAAGTGAAGA 810
|::: |||||:::
QY 1319 uLeu 1320
|::: |||||:::
Db 811 ACTG 814

RESULT 5
BQ701900
LOCUS
DEFINITION BQ701900 595 bp mRNA linear EST 11-DEC-2001
laevis cDNA clone XL095nl8 5', mRNA sequence.
BQ701900
VERSION BQ701900.1 GI:17502089
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 595)
REFERENCE Kiteyama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
AUTHORS Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
```

```

Email: tshini@genes.nig.ac.jp.
FEATURES
  source
    1. .595
      /location="Qualifiers"
      /db_xref="taxon:8355"
      /clone="X1093n18"
      /clone_lib="NIBB Mochii normalized Xenopus tailbud
      library"
      /tissue_type="whole embryo"
      /dev_stage="stage 25"
      /note="vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
      were oligo-dT primed and directionally cloned. Staging
      according to Nieuwkoop and Faber. Library is subtracted
      and was constructed by N. Garrett and A.M. Zorn,
      (Wellcome/CRC Institute)."
BASE COUNT      186 a   133 c   140 g   136 t
ORIGIN
Alignment Scores:
Pred. No.:      9.52e-72      Length:      595
Score:          908.00      Matches:      164
Percent Similarity: 91.37%      Conservative: 16
Best Local Similarity: 83.25%      Mismatches: 17
Query Match:    13.18%      Indels:      0
DB:             13          Gaps:      0

US-09-818-990B-2 (1-1320) x BJ071900 (1-595)
Qy 1062 AspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAsp 1081
Db 5 GAGAGCGAGATAGTAGAGGAGCGATTCGCGACCATATCTTCAGCGCTCCAGGAGAT 64
Qy 1082 MetValAlaHisGluGlyArgLeuGlyCysArgLeuAspCysLysValSerGlyLeuProPro 1101
Db 65 ATGGTTGCACACGAGAGCGACTCGCGACTCGATTGTAGGTAGCGGCTCCCAACT 124
Qy 1102 ProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMet 1121
Db 125 CCAGATATTATGTGGCTACTCAATGGAAGGCAAGTAAACCCAGATGCTGTCCACAAATG 184
Qy 1122 LeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArgAspAla 1141
Db 185 CTTGTGAGGAGAAATGGAATTCATTCGCTCTTATTGATCCTTAACACAGCAGATCT 244
Qy 1142 GlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeu 1161
Db 245 GGGACCTACACTTCGATTCGCACTTAATAAACCGGGCAAACTCATTCAGCTTGAGCTC 304
Qy 1162 SerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys 1181
Db 305 AGTATCATGGCAAGAAGTCCAAAAGCACCAGTATTTTGGAAAACTTCAAAATGT 364
Qy 1182 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProPro 1201
Db 365 GGCCTCCAGAGGCCATCCAGTCAGATTAGAGAGTCGAGTTGTGGAATGCCCCACCA 424
Qy 1202 ValPheTrpTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet 1221
Db 425 GTGTTCTATTGGGAAGAAGACACACATCCATCCGCCCAATAGACACAGGATGAGCATG 484
Qy 1222 HisGlnAspThrThrGlyTrpAlaCysLeuLeuIleGlnProAlaLysLysSerAspAla 1241
Db 485 CATCAAGATAACACTGGTTAGTGCTTCTCATCCAGCCTGTAGAAAAAGAGATGCA 544
Qy 1242 GlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThr 1258
Db 545 GGATGGTACAGCTGTGACGAAAGAAATGAAGCAGGAGATTGTATCTGCACG 595

RESULT 6
LOCUS      BI691193      943 bp      mRNA      linear      EST 18-SEP-2001
DEFINITION 60314585F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5354589 5',
mRNA sequence.

```

```

BI691193      GI:15653822
VERSION      BI691193.1
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 943)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Jeffrey Green M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1AM11901 row: i column: 22
              High quality sequence stop: 809.
              Location/Qualifiers
                source
                  1. .943
                    /organism="Mus musculus"
                    /strain="FVB/N"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:5354589"
                    /clone_lib="NCI_CGAP_Mam6"
                    /sex="female, virgin"
                    /tissue_type="infiltrating ductal carcinoma"
                    /dev_stage="5 months"
                    /lab_host="DH103"
                    /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                    Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                    Library constructed by Life Technologies. Investigator
                    providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT   245 a   256 c   250 g   192 t
ORIGIN
Alignment Scores:
Pred. No.:      1.54e-69      Length:      943
Score:          887.00      Matches:      169
Percent Similarity: 73.75%      Conservative: 53
Best Local Similarity: 56.15%      Mismatches: 69
Query Match:    12.87%      Indels:      10
DB:             13          Gaps:      2

US-09-818-990B-2 (1-1320) x BI691193 (1-943)
Qy 930 AspGluIleGlnHisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLys 949
Db 2 GACGAGTGGAGGATCCAGATGCTCTGGAGAACCAACAGCTCCTCTTTGAGATG 61
Qy 950 ArgLeuLysHisPheArgValThrGluGlySerProValThrPheThrCysLysIleVal 969
Db 62 AAGCTGAACACACTACAAGATCTTTGAGGGGATCGCGTGACTTTCACGTGTCGAGTGGCT 121
Qy 970 GlyIleProValProLysValTrpPheLysAspGlyLysGlnIleSerLysArgAsn 989
Db 122 GGAATCAAGGCCAAAGGATCTATTGTTTAAAGATGGGAGCAGATTTCTCCGAAGAGC 181
Qy 990 GluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThr 1009
Db 182 GATCCTACACCATTCAGAGAGACCTTGTATGGGACCTGCTCTCTCCACACACCGCCTCT 241
Qy 1010 ThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSer 1029
Db 242 ACCCTAGACGAGATGGAACTACACCATCATGCTGCCAACCTCAGGTCGGTCCAGT 301
Qy 1030 CysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArg-----LeuThrSer 1047
Db 302 TGTACAGGAAGGCTTAATGGTACAGGCTGTCAACCAAGAGCGCGCGCTCTCC 361

```



```
QY 1048 AlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgasp-----Lys 1063
      :::||| ||| ||| ::: |||||
Db 362 TCAGGCCATCTCATGCCAAGAGCCCTCGCTCGATCAGCGGACAGTGGAGATGAAC 421
      :::||| ||| ||| ::: |||||

QY 1064 GluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetVal 1083
      :::||| ||| ||| ||| ::: |||||
Db 422 GAGCCCATTCAGGAGCGATTCTTCAGACCTCACTTCCTGCGAGGCTCTGGAGACCTGACC 481
      :::||| ||| ||| ||| ::: |||||

QY 1084 AlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeuProProGlu 1103
      :::||| ||| ||| ||| ::: |||||
Db 482 GTTCAGGAAGGCAAGCTCTCGAGATGGATGCAAGGTCAAGTGGTTACCAACCCAGAT 541
      :::||| ||| ||| ||| ::: |||||

QY 1104 LeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuVal 1123
      :::||| ||| ||| ||| ::: |||||
Db 542 CTCAGTGGCACTAGATGAAGCCCATACGCCCGACAGTGTCTCACAGATGCTGTGTC 601
      :::||| ||| ||| ||| ::: |||||

QY 1124 ArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThr 1143
      :::||| ||| ||| ||| ::: |||||
Db 602 CGTGAGAATGGGTGCCACTCCCTCATATAGAGCCAGTCACTCCCGGAGCGCGGCATC 661
      :::||| ||| ||| ||| ::: |||||

QY 1144 Tyr-LysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGlu-LeuSerV 1163
      :::||| ||| ||| ||| ::: |||||
Db 662 TACACATGTATTGCCACCAACAGCAGCAGACAGACTCGTTTAACTGGAGCTGTGTGG 721
      :::||| ||| ||| ||| ::: |||||

QY 1163 alValAlaLys-GluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGly 1182
      :::||| ||| ||| ||| ::: |||||
Db 722 TTGCTGCTAGAGAGACACACAAATGCCCTGTGTATTATGAGAGACTACGAACACGGGG 781
      :::||| ||| ||| ||| ::: |||||

QY 1183 ValProGluGlyHisProVal-ArgLeuGluCysArgValIleGlyMetProProProVa 1202
      :::||| ||| ||| ||| ::: |||||
Db 782 TTGCTGATTGGATACCCAGTGGGTGGGAATGCGTCTCCGGGAATGCCCGCACCTCA 841
      :::||| ||| ||| ||| ::: |||||

QY 1202 lPheTyTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSer 1220
      :::||| ||| ||| ||| ::: |||||
Db 842 GATTATTGGAGAAGAACCACTGATCGGTCACTCAAAAGCACTGAGCAAGTAACA 896
      :::||| ||| ||| ||| ::: |||||

RESULT 7
BG827059 868 bp mRNA linear EST 22-MAY-2001
LOCUS BG827059
DEFINITION 602749109F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4901666 5',
      mRNA sequence.
ACCESSION BG827059
VERSION BG827059.1 GI:14174646
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 868)
      NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapb-f@mail.nih.gov
      Tissue Procurement: ATCC
      cDNA Library Preparation: Ling Hong/Rubin Laboratory
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone Distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LNL at:
      http://image.llnl.gov
      Plate: LCM1797 row: b column: 03
      High quality sequence stop: 807.
      Location/Qualifiers
          1..868
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:4901666"
             /clone_lib="NIH_MGC_17"
             /tissue_type="rhabdomyosarcoma"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
```

```
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
BASE COUNT 262 a 200 c 219 g 187 t
ORIGIN
Alignment Scores:
Pred. No.: 1.08e-68 Length: 868
Score: 877.00 Matches: 169
Percent Similarity: 77.62% Conservatives: 46
Best Local Similarity: 61.01% Mismatches: 52
Query Match: 12.73% Indels: 11
DB: 12 Gaps: 2
US-09-818-990b-2 (1-1320) x BG827059 (1-868)
QY 901 GlnGluTyrlsSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu 920
      :::||| ||| ||| ||| ::: |||||
Db 21 CAGGAATACAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAGATACAGGCTA 80
      :::||| ||| ||| ||| ::: |||||

QY 921 GluArgThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGly 940
      :::||| ||| ||| ||| ::: |||||
Db 81 GAAAGGTCTCTGTGGATGAATCAGGTGATCAAGTTCAGTTCAGAGATGTCCTGTGAA 140
      :::||| ||| ||| ||| ::: |||||

QY 941 LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer 960
      :::||| ||| ||| ||| ::: |||||
Db 141 AATGGAATGCAACCATCTTTTGGATGAAGTGAACATTTACAAGATCTTTGAGGGGAA 200
      :::||| ||| ||| ||| ::: |||||

QY 961 ProValThrPheThrCysLysIleValGlyIleProValProLysValTyTrpPheLys 980
      :::||| ||| ||| ||| ::: |||||
Db 201 CCAGTAATTTTCATGTAGATGGCTGGAAATCCAAAGCCAAAGATCTATTTGTTTAA 260
      :::||| ||| ||| ||| ::: |||||

QY 981 AspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluGlyAspGly 1000
      :::||| ||| ||| ||| ::: |||||
Db 261 GATGGGAAGCAGATCTCTCCAAAGAGTGCATCCTACACCATTCAAAGAGATCTCGATGG 320
      :::||| ||| ||| ||| ::: |||||

QY 1001 ThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyThrIleMet 1020
      :::||| ||| ||| ||| ::: |||||
Db 321 ACCTGCTCCCTCCATACACAGCCTCCACCTAGATGATGGGAATTTATACAATTATG 380
      :::||| ||| ||| ||| ::: |||||

QY 1021 AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro 1040
      :::||| ||| ||| ||| ::: |||||
Db 381 GCTCAAAACCTTCAGGGCCGCATCAGTTGTACTGGACGGCTAATGGTACAGGCTGTCAAC 440
      :::||| ||| ||| ||| ::: |||||

QY 1041 IleArg-SerArgLeuThrSerAlaGlyGlnSerHis-----ArgG1 1054
      :::||| ||| ||| ||| ::: |||||
Db 441 CAAGACGGTGAAGTCCCGGTCTCCCTCAGGCCATCCTCATGTCAGAAGCCCTCGTTC 500
      :::||| ||| ||| ||| ::: |||||

QY 1054 yArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArgProHI 1074
      :::||| ||| ||| ||| ::: |||||
Db 501 TAGATCAAGGGACAGTGGAGACGAAAAATGAACCAATTCAGGAGCGATTTCTTCAGACCTCA 560
      :::||| ||| ||| ||| ::: |||||

QY 1074 sPheLeuAlaProGlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCy 1094
      :::||| ||| ||| ||| ::: |||||
Db 561 CTTCTTCAGGCTCCTGGAGATCTGACGTTCAGAAAGGAAACCTCTCAGAAATGGAGCTG 620
      :::||| ||| ||| ||| ::: |||||

QY 1094 sLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLe 1114
      :::||| ||| ||| ||| ::: |||||
Db 621 CAAAGTCAGTGGGTACCAACCCAGATCTAAGCTGGCAACTAGATGAAGACCCGCTACG 680
      :::||| ||| ||| ||| ::: |||||

QY 1114 uProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis---SerLeuLeuI1 1133
      :::||| ||| ||| ||| ::: |||||
Db 681 CCCTGACAGTGTCTCAAGATGCTGGTGGTGAACGGGTGCACTACTCTGCG-ATCAT 739
      :::||| ||| ||| ||| ::: |||||

QY 1133 eAspProLeu-ThrGlnArgAspAlaGlyThrTyTrpLysCysIle-AlaThrAsnLysThr 1152
      :::||| ||| ||| ||| ::: |||||
Db 740 AGAGCCAGTCCACGTCACGTGATCCCGCATCTACACATGATTATTAGCTACCAAGCGACA 799
      :::||| ||| ||| ||| ::: |||||

QY 1153 GlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGlu 1167
```


The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized cartilaginous tumor library cDNA library preparation: M.B. Soares lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES		Location/Qualifiers			
source					
1. .717		/organism="Rattus norvegicus"			
/strain="Sprague-Dawley"		/db_xref="taxon:10116"			
/clone="UI-R-D21-cm-k-05-0-UI"		/clone_lib="UI-R-D21"			
/tissue_type="Chondrosarcoma"		/dev_stage="37 days"			
/lab_host="DH10B (Life Technologies)"		/notes="Organ: Spine; Vector: pT7n3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-D21 is a normalized cDNA library containing the following tissue(s): Swarm Rat Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7n3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CATCTCTGTA. The Rat cartilaginous tumor tissue was provided by Dr Jeff Stevens at the University of Iowa.			
TAG_LIB=UI-R-D21		TAG_TISSUE=cartilaginous tumor			
TAG_SEQ=CATTCTGTGA					
BASE COUNT	129 a 188 c 203 g 196 t	1 others			
ORIGIN					
Alignment Scores:					
Pred. No.:	3.8e-67	Length:	717		
Score:	858.50	Matches:	154		
Percent Similarity:	79.15%	Conservative:	32		
Best Local Similarity:	65.53%	Mismatches:	42		
Query Match:	12.46%	Indels:	7		
DB:	14	Gaps:	2		
US-09-818-990B-2 (1-1320) x BQ206762 (1-717)					
Qy	1049	GlyGlnSerHisArgGlyArgSerArgValGlnGluArgasp-----LysGlu	1064		
Db	712	GGTCATCTCTCAGCAGAGCCCGCTCTCGATCAGCGACAGTGGAGTGAATGAG	653		
Qy	1065	ProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetValala	1084		
Db	652	CCCATTCAGAGCGATCTTCAGACCTCACTCTCTCGAGGCTCTGGAGACCTCAGCGTT	593		
Qy	1085	HisGluGlyArgLeuGlyArgLeuAspCysLysValSerGlyLeuProProGluLeu	1104		
Db	592	CAGGAAGGCAAGCTCTCGAGATGAGTGCAGAGTCAAGTTCAGTGGATTACCAACCCAGATCTC	533		
Qy	1105	ThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValarg	1124		
Db	532	AGCTGGCAACTAGACGGAAGACCCATCCGCCCTGACAGTGTCTCAAGATCTGTCGGT	473		
Qy	1125	GluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyr	1144		
Db	472	GAGATGGGTGGCACTCCCTCATTTATAGAGCCAGTCACTCCCGGAGGAGGATCTAC	413		
Qy	1145	LysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValVal	1164		

Db	412	ACCTGCATCGCCACCAACAGAGCAGCGCCAGAACTCATTTAACCTGGAGCTTGCTGCT	353
Qy	1165	AlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValPro	1184
Db	352	GCTAAAGACACACAGGCCCTGTGTATCGAAGCTGCAAAACACGGGGTGGCC	293
Qy	1185	GluGlyHisProValArgLeuGluCysArgValIleGlyMetProProValPheTyr	1204
Db	292	GATGGGTACCCCGTGGCTGGAATCGCGGTTCGGGGTGGCCGCCACCTCAGATATTC	233
Qy	1205	TrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAsp	1224
Db	232	TGAAGAAAGAAAATGAATCGCTCACTCAGCAGCACTGATCGAGTGCAGCATCCAGGAT	173
Qy	1225	ThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyr	1244
Db	172	ATCATGGCTACATCTGCCTGCTCATTCAGGGAGCACAAAGAGAGCGTGGGTGGTAT	113
Qy	1245	ThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyr	1264
Db	112	ACCGTCCCGCAAGAGCAGGAGCATCGTGTCTGTACTGCCAGCTGGATGTCTAC	53
Qy	1265	AlaGlnTrpHisHisGlnIleProProMetSerValArgPro	1279
Db	52	ACCAGTGGCATCAGCAG-----CCACAGACCACCAAGCCA	17
RESULT 10			
BF126187			
LOCUS	BF126187	553 bp	mRNA linear EST 24-OCT-2000
DEFINITION	601650450F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934045 5', mRNA sequence.		
ACCESSION	BF126187		
VERSION	BF126187.1	GI:10965227	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 553)		
TITLE	NIH-MGC http://mgi.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: coqabs-remail.nih.gov		
	Tissue Procurement: CLONETECH Laboratories, Inc.		
	cDNA Library Preparation: CLONETECH Laboratories, Inc.		
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LUCM774 row: d column: 14		
	High quality sequence stop: 542.		
FEATURES	Location/Qualifiers		
source	1. .553		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3934045"		
	/clone_lib="NIH_MGC_76"		
	/lab_host="DH10B (TI phage-resistant)"		
	/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCCGAGCGGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."		
BASE COUNT	143 a 138 c 153 g 119 t		
ORIGIN			

Alignment Scores:
Pred. NO.: 9.71e-66 Length: 553
Score: 841.00 Matches: 158
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.21% Indels: 0
DB: 12 Gaps: 0

US-09-818-990B-2 (1-1320) x BF126187 (1-553)

```
Qy 1163 ValValAlaLysGluValLysLysAlaProValLleLeuGluLysLeuGlnAsnCysGly 1182
Db 2 GTAGTAGCCAAAGAGGTGAAGAAACACCTGTGCTGGGAAACTACAGAACTGCGGT 61
Qy 1183 ValProGluGlyHisProValArgLeuGluCysArgValLleGlyMetProProVal 1202
Db 62 GTTCCCGAAGCCACCCGCTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACCTGTG 121
Qy 1203 PheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMethis 1222
Db 122 TTCTACTGGAAGAAGACAATGAGACCATCCTTGCACCAGAGAGAGATCAGTATGCAC 181
Qy 1223 GlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGly 1242
Db 182 CAGGACAAACAGGGTATGCTGCTTCTCATTCAGCCAGCAAGAAATCAGACGCTGGA 241
Qy 1243 TrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 1262
Db 242 TGGTACAGTTGTGAGCCCAAGAAAGTGAAGCCGGCATCGTGTGTCACATGCCAGGCTGGAT 301
Qy 1263 IleTyrAlaGlnTrpHisHisGlnIleProProMetSerValArgProSerGlySer 1282
Db 302 ATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGTCTGTCGGCCAGTGGCAGT 361
Qy 1283 ArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerMetGlu 1302
Db 362 CGCTACGGATCTTCACACAGTAAAGGACTTGACATATTTTTCGCTTTTCCTCCATGGA 421
Qy 1303 SerThrMetValTyrSerCysSerSerArgSerValValGluSerAspGluLeu 1320
Db 422 AGCAGATGGTGTATTCATGCTCTCTCGGAGTGTAGTGGAGAGTGAAGT 475
```

RESULT 11
BI526367 602925663F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5058231 5',
LOCUS mRNA sequence.
DEFINITION NIH-MGC http://mgi.nci.nih.gov/.
ACCESSION BI526367.1 GI:15351159
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 828)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1157 row: m column: 16
High quality sequence start: 25
High quality sequence stop: 826.
FEATURES
source Location/Qualifiers
1..828

```
/organism="Mus musculus"  
/strain="CZECH II"  
/db_xref="taxon:10090"  
/clone="IMAGE:5058231"  
/clone_lib="NCI_CGAP_Lu33"  
/tissue_type="pooled lung tumors"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a  
modified polylinker; Site:1: NotI; Site:2: EcoRI; 1st  
strand cDNA was prepared from mRNA obtained from pooled  
lung tumors with a NotI - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGGCGCTCTGTTTCTTTTCTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with NotI and cloned into the Not  
I and Eco RI sites of the modified p7T3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 211 a 228 c 223 g 166 t  
ORIGIN
```

Alignment Scores:

Pred. No.: 1.45e-64 Length: 828
Score: 831.00 Matches: 159
Percent Similarity: 76.21% Conservative: 46
Best Local Similarity: 59.11% Mismatches: 53
Query Match: 12.06% Indels: 11
DB: 13 Gaps: 2

US-09-818-990B-2 (1-1320) x BI526367 (1-828)

```
Qy 931 GluIleGlnHisAspGluIleProThrGlyLysCys-IleAlaProIlePheAspLysar 950
Db 30 AAGTGCAGGATCCAGATGTGCTGTGGAGAACGCCAACACCTCCCTCTTTGAGATGAA 89
Qy 950 gLeu-LysHisPheArgValThrGluGlySerProValThrPheThrCysLysIleValg 970
Db 90 GCTGGAACACTACAGATCTTTGAGGGGATGCGGGTACTTTCACGCTGTCGAGTGGCTG 149
Qy 970 lyIleProValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnG 990
Db 150 GGATCCAAAGCCAAAGATCTATTGGTTAAAGATGGAGACAGATTTCTCCGAGAGCG 209
Qy 990 luHisCysLysMetArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThr 1010
Db 210 ATCACTACCATTCAGAGACCTTGATGGACCTGCTCTCCACACACCGCTCTCTA 269
Qy 1010 hrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerC 1030
Db 270 CCCTAGACGAGTGGGAACATACACCATCATGTGCTGCCAACCCCTCAGGGTCGCTCAGTT 329
Qy 1030 ysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArg-----LeuThrSerA 1048
Db 330 GTACAGGAAGCTAATGGTACAGGCTGTCAACCAAGAGCGCGGAGTCCCGCTCTCCCT 389
Qy 1048 lagIleGlnSerHisArgGlyArgSerArgValGlnGluArgAsp-----LysG 1064
Db 390 CAGGCCATCTCATGCCAAGAGCGCTGCTCTCATGATCAGCGGACAGTGGAGATGAAACG 449
Qy 1064 luProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetVala 1084
Db 450 AGCCCATTCAGAGCGGATTTCTCAGACCTCACTTCTCAGGCTCCTGGAGACCTGACCG 509
Qy 1084 laHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluL 1104
Db 510 TTCAGGAAGGCAAGCTCTGAGGATGGACTCAAGGTGATGGGTTCACCAACCCAGATC 569
Qy 1104 euThrTrpLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuVala 1124
Db 570 TCAGCTGGCAACTAGATGGAAAGCCCATACGCCCGCAGAGTCTCACAAAGATGTGTGCC 629
Qy 1124 rgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspIleGlyThr 1144
Db 630 GTGGAATGGGGTCCCTCACTTATTAGACGAGTACAGTCCCGGACCGCGGCATCT 689
```

US-09-818-990B-2 (1-1320) x BM795050 (1-612)

[illegible]

was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 157 a 192 c 177 g 128 t

ORIGIN

Alignment Scores:

Pred. No.: 5,02e-59 Length: 654
Score: 768.00 Matches: 140
Percent Similarity: 79.36% Conservative: 33
Best Local Similarity: 64.22% Mismatches: 39
Query Match: 11.15% Indels: 6
DB: 10 Gaps: 2

US-09-818-990B-2 (1-1320) x BB548628 (1-654)

QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
Db 1 ATTCAGAGAGACCTTGATGGAGACCTGCTCTCCACACACCGCCCTACCCCTAGAGAGC 60
QY 1014 AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033
Db 61 GATGGGAACCTACACCATCATGCTGCCAACCCCTCAGGTCGCTCAGTGTCTACAGGAGG 120
QY 1034 LeuMetValGlnSerLeuProIleArgSerArg-----LeuThrSerAlaGlyGlnSer 1051
Db 121 CTAATGGTACAGCTGTCAACCAAGAGCGCCAGCTCCCGCTCTCCTCAGGCCATCCT 180
QY 1052 HisArgGlyArgSerArgValGlnGluArgAsp-----LysGluProLeuGln 1067
Db 181 CATGCCAAGAGCGCTCGCTCTCGATCAGCGGAGTGGAGATGAAACGAGCCCATTCAG 240
QY 1068 GluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGly 1087
Db 241 GACGGATTCTTACAGACCTCACTTCCTCGACGCTCTCTGGAGACCTGACCGTTCAGGAAGC 300
QY 1088 ArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuLeuThrTrpLeu 1107
Db 301 AAGCTCTGCAGGATGGACTCAAGGTCAGTGGTGTACCAACCCAGATCTCAGCTGGCAA 360
QY 1108 LeuAsnGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGly 1127
Db 361 CTAGATGGAAGCCCATACCCCGCAGTGCTCACAAGATGCTGGTCCGTGAGATGGG 420
QY 1128 ValHisSerLeuLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTyLysCysIle 1147
Db 421 GTCCACTCCCTCATATAGACCACTCAGTCCCGGACCGCGCATCTACATGTATT 480
QY 1148 AlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGlu 1167
Db 481 GCCACCAACAGACGAGACAGAACTCGTTTAACCTGGAGCTTGTGTGCTTAAGGAA 540
QY 1168 ValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHis 1187
Db 541 GCACACAGCCCTCTGTGTTATGGAGAGCTACAGAACACGGGGTGTGTGATGGATAC 600
QY 1188 ProValArgLeuGluCysArgValIleGlyMetProProProValPheTyTrp 1205
Db 601 CCAGTGGCGTGAATGCGCTGCTCGGAGTGCCGCCACCTCAGATATTG 654

RESULT 15 BG923162 772 bp mRNA linear EST 05-JUN-2001
LOCUS 602824001F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4952713 5',
DEFINITION mRNA sequence.
ACCESSION BG923162
VERSION BG923162.1 GI:14303638
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 772)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10911 row: a column: 02
High quality sequence stop: 765.

FEATURES
source

1. .772
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:4952713"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
Providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 198 a 229 c 202 g 143 t
ORIGIN

Alignment Scores:

Pred. No.: 5,2e-58 Length: 772
Score: 758.00 Matches: 140
Percent Similarity: 80.48% Conservative: 29
Best Local Similarity: 66.67% Mismatches: 38
Query Match: 11.00% Indels: 3
DB: 13 Gaps: 1

US-09-818-990B-2 (1-1320) x BG923162 (1-772)

QY 1092 LeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGln 1111
Db 1 ATGGACTGCAAGTCTAGTGGTTACCAACCCAGATCTCAGTGGCACTAGATGGAAG 60
QY 1112 ProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1131
Db 61 CCCATACGCCCGACAGTGTCTACAAAGATGCTGCTGCTGAGATGGGTCCACTCCCTC 120
QY 1132 LeuIleAspProLeuThrGlnArgAspAlaGlyThrTyLysCysIleAlaThrAsnLys 1151
Db 121 ATTATAGCCAGTCACTCCCGGACCGCGGCATCTACATGTATGCCACCAACAGA 180
QY 1152 ThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysAla 1171
Db 181 GCAGCAGACAACTCGTTTAACCTGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 1172 ProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeu 1191
Db 241 CCTGTGTTTATGGAGAAGCTACAGAACACGGGGTGTCTGATGATACCCAGTGGCGTG 300
QY 1192 GluCysArgValIleGlyMetProProValPheTyTrpLysLysAspAsnGluThr 1211
Db 301 GAATGCGGTGCTCTCGGAGTGGCCCGCCACCTCAGATATTGGAAGAAGAAATGAATCG 360
QY 1212 IleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyAlaCysLeu 1231
Db 361 CTCACCTCACGACCTGAGCGAGTAAAGCATGCACAGGATAATATGCTGCTACATCTGCTG 420
QY 1232 LeuIleGlnProAlaLysLysSerAspAla-GlyTyTrpThrLeuSerAlaLysAsnGln 1251
Db 421 CTCATCCAGGAGCCACAAAGAGACGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480

Qy 1251 uAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis----- 1269
Db 481 AGCAGGCATTGTCTGCTGCACTGCCCGGCTGGATGCTACACCCAGTGGCACCAGCAGCC 540
Qy 1270 -GlnIleProProMetSerValArgProSerGlySerArgTyrGlySerLeuThrSe 1289
Db 541 ACAGACCACCAAGCCAAAAAAGTACGGCCCTCGGCCAGTCGCTACGCAGCACTTTCGGA 600
Qy ~ 1289 rLysGlyLeuAspIlePheSerAlaPhe 1298
Db 601 CCAGGGACTAGACATCAAAAGCCGCTTTC 628

Search completed: November 30, 2002, 21:32:34
Job time : 2671 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 14:01:04 ; Search time 6626 Seconds
(without alignments)
17406.342 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 3963

Sequence: 1 atgcaagacagacataga.....tggagagtgtgaacttttaa 3963

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3504	88.4	5707	9	AF328296	Homo sapi
2	3453	87.1	5804	9	HS805275	Homo sapi
3	3249	82.0	5486	9	HS803309	Homo sapi
4	2602	65.7	4683	9	HS803687	Homo sapi
5	1483	37.4	2127	9	AK027343	Homo sapi
6	904	22.8	160558	9	AC024258	Homo sapi
7	904	22.8	175265	2	AC063964	Homo sapi
8	904	22.8	180707	2	AC024460	Homo sapi
9	413	10.4	11170	9	AL512429	Human DNA
10	413	10.4	290625	2	AL442068	Homo sapi
11	236	6.0	175265	2	AC063964	Homo sapi
12	227	5.7	290625	2	AL442068	Homo sapi
13	210	5.3	166869	9	AC016395	Homo sapi
14	41	1.0	151002	2	AC114220	Rattus no
15	38	1.0	151002	2	AC114220	Rattus no
16	37	0.9	512	11	AF139038	Ovis arie
17	26	0.7	42861	2	AC120168	Mus muscu
18	24	0.6	3135	3	AY057969	Girardia
19	23	0.6	71656	2	AC067914	Homo sapi
20	23	0.6	128824	2	AC094957	Rattus no
21	23	0.6	147603	9	HS025883	Human DNA
22	23	0.6	173222	9	AC009127	Homo sapi
23	23	0.6	194732	9	AB042031	Homo sapi
24	23	0.6	195741	2	AC098273	Rattus no
25	23	0.6	300000	9	AP002529	Homo sapi
26	22	0.6	68743	2	AC113547	Mus muscu
27	22	0.6	136493	2	AC125500	Takifu
28	22	0.6	170000	2	AC004394	Homo sapi
29	22	0.6	170000	2	AC004579	Homo sapi
30	22	0.6	174138	9	AC008871	Homo sapi
31	22	0.6	174891	2	AC113847	Rattus no
32	22	0.6	190000	2	AC004580	Homo sapi
33	22	0.6	196426	2	AC127234	Mus muscu
34	22	0.6	210515	9	AC097382	Homo sapi
35	21	0.5	1580	10	AF205078	Mus muscu
36	21	0.5	3147	10	BC027364	Mus muscu
37	21	0.5	37443	9	U73627	Human Chrom
38	21	0.5	51761	2	AC068574	Homo sapi
39	21	0.5	62726	2	AC084090	Homo sapi
40	21	0.5	70406	2	AC103376	Mus muscu
41	21	0.5	71530	2	AC116112	Mus muscu
42	21	0.5	94851	9	AC004923	Homo sapi
43	21	0.5	94911	8	AT114E10	Arabidops
44	21	0.5	106922	2	AC127029	Homo sapi
45	21	0.5	108816	2	AL670379	Homo sapi
46	21	0.5	115279	9	AC112506	Homo sapi
47	21	0.5	127188	9	AC010490	Homo sapi
48	21	0.5	127539	9	CNS01DSN	Human chr
49	21	0.5	132830	2	AC011051	Homo sapi
50	21	0.5	138915	2	AP005246	Oryza sat
51	21	0.5	142918	2	AF252828	Homo sapi
52	21	0.5	146558	2	AC090753	Homo sapi
53	21	0.5	152141	2	AC119787	Rattus no
54	21	0.5	153241	2	AC109690	Rattus no
55	21	0.5	156307	2	AP005248	Oryza sat
56	21	0.5	162143	10	AL596283	Mouse DNA
57	21	0.5	162473	9	AC005803	Homo sapi
58	21	0.5	167875	9	AC083906	Homo sapi
59	21	0.5	170656	2	AF216672	Homo sapi
60	21	0.5	177654	2	AC025988	Homo sapi
61	21	0.5	178229	2	AC009420	Homo sapi
62	21	0.5	185045	2	AC126303	Rattus no
63	21	0.5	185148	9	AC096641	Homo sapi
64	21	0.5	186824	2	AC121881	Mus muscu
65	21	0.5	200833	9	AL390738	Human DNA

c 942	18	0.5	496	8	OAL410059
c 943	18	0.5	497	8	OJA410062
c 944	18	0.5	497	8	OJA410065
c 945	18	0.5	501	6	AX433981
c 946	18	0.5	504	8	AF402443
c 947	18	0.5	506	8	AF402442
c 948	18	0.5	513	5	ABY13266
c 949	18	0.5	515	5	ABAJ823
c 950	18	0.5	533	11	AU025751
c 951	18	0.5	534	8	CAF236981
c 952	18	0.5	540	8	CAF236982
c 953	18	0.5	580	8	ABU78843
c 954	18	0.5	606	8	AF129480
c 955	18	0.5	635	9	HS4322986
c 956	18	0.5	660	8	CNS01AI3
c 957	18	0.5	695	9	HS4336992
c 958	18	0.5	711	14	AF485891
c 959	18	0.5	711	14	PVXCPG
c 960	18	0.5	711	14	PVXCPK
c 961	18	0.5	711	14	PVXCPNL1
c 962	18	0.5	711	14	PVXCPXA
c 963	18	0.5	712	6	AX342361
c 964	18	0.5	714	14	AF202462
c 965	18	0.5	714	14	AF202462
c 966	18	0.5	720	14	AF528555
c 967	18	0.5	720	8	CNS01D6W
c 968	18	0.5	731	14	BR5ATTGLYA
c 969	18	0.5	731	14	BR5ATTGLYB
c 970	18	0.5	731	14	BR5ATTGLYC
c 971	18	0.5	731	14	BR5ATTGLYD
c 972	18	0.5	731	14	BR5ATTGLYE
c 973	18	0.5	731	14	BR5ATTGLYG
c 974	18	0.5	750	14	AF260640
c 975	18	0.5	750	14	AF260641
c 976	18	0.5	756	14	BR242713
c 977	18	0.5	756	14	BR242714
c 978	18	0.5	756	14	BR242715
c 979	18	0.5	759	9	HS4325398
c 980	18	0.5	764	9	HS4337632
c 981	18	0.5	768	3	AF171047
c 982	18	0.5	774	14	BR5GATTPRO
c 983	18	0.5	774	14	BR5GGLYPRO
c 984	18	0.5	774	14	BR333539
c 985	18	0.5	789	14	POX505748
c 986	18	0.5	789	14	PVU19790
c 987	18	0.5	789	14	PVXCP
c 988	18	0.5	795	6	E24616
c 989	18	0.5	818	6	AX004604
c 990	18	0.5	826	14	BR5V22069
c 991	18	0.5	838	14	BR5GGLYQ
c 992	18	0.5	840	6	AX047076
c 993	18	0.5	861	6	I19731
c 994	18	0.5	861	9	HUMDSPHS
c 995	18	0.5	868	9	HS4337312
c 996	18	0.5	900	6	AX028276
c 997	18	0.5	950	9	HS4337496
c 998	18	0.5	972	8	ANNMP
c 999	18	0.5	974	9	HS4337350
c1000	18	0.5	993	9	BC004960

ALIGNMENTS

RESULT 1	AF328296	Homo sapiens	5707 bp	mRNA	linear	PRI 06-MAY-2001
LOCUS	AF328296	Homo sapiens	5707 bp	mRNA	complete cds.	
DEFINITION	AF328296	Homo sapiens	5707 bp	mRNA	complete cds.	
ACCESSION	AF328296	Homo sapiens	5707 bp	mRNA	complete cds.	
VERSION	AF328296.1	Homo sapiens	5707 bp	mRNA	complete cds.	
KEYWORDS						
SOURCE						
ORGANISM						

REFERENCE	1 (bases 1 to 5707)
AUTHORS	Bang, M.L., Mudry, R.E., McEllinny, A.S., Trombitas, K., Geach, A.J., Yamasaki, R., Sorimachi, H., Granzier, H., Gregorio, C.C. and Labat, S.
TITLE	Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in z-disc and I-band protein assemblies
JOURNAL	J. Cell Biol. 153 (2), 413-427 (2001)
MEDLINE	21206024
PUBMED	11309420
REFERENCE	2 (bases 1 to 5707)
AUTHORS	Bang, M.-L.
TITLE	Direct Submission
JOURNAL	Submitted (13-DEC-2000) Structural and Computational Biology, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg 69117, Germany
FEATURES	Location/Qualifiers
source	1. 5707
CDS	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="10"
	/map="10q21.1"
	/tissue_type="striated muscle"
	488..4450
	/codon_start=1
	/product="myopalladin"
	/protein_id="AAK50625.1"
	/db_xref="GI:13957727"
	/translation="MODDSIEASTSISQLLRSEYLAETHRRGNRNSRAEPSNPCHFGSPGAEGGGDDLPDLFSLAQSEELDESYNLARLAIYDPLEKADTQARKRLSPDKHSPNLSFEPFQDNPR4PTSKSPQAKRPQCYSETSKKVFUNKADFTEE LSGLFKSHSKRIRPRACKNHSKLSQNKVQENSSFDLSERERSVPTIIPAD TRDNEVHALAQEQAARQAAGSDTTCGSSPSLYEELPGOPPTOKLR SREVPCTRVQLDCIVGIPPPQVRYCEGKLENSPDHIVQAGNLHSLTTAEAFEE DRYSCFASNIYDSTSIETGVSDDSEDPNKEEMRNIRKQNPNEVSPPTTSA VLPVPAQHLVAQPRVATIQCCQPTNYLQGLDGKPTIAAPVFTKMLQNLASSEG LVVFECEVAGAPSPKVEWREGTLIEDSPDFRILQKKPRMAEPETICTLVIAEVAE DSGCFTCTASNYGVSSIAQLHVRGNEDSLNNGLSHSANSPTNLAAIPQSPPHSE PPSVEOPPKLEGVLVNHNEPRSSRIGLVRHFNLPEDDKGSEASEAGVVTROT RPSDFRFGNQATKTPERSPPVKPPVLAKPKLDSTOQLQHNQVLLHQHQLONPP SPKPEPFNNTVLNSNAPVATSNKQVAPSKQSGSLQNEPLPPGTEPTPTTPTTSPFTLSTPQTQRTSLELLVSHPSVQTKSGGLSLQNEPLPPGTEPTPTTPTTPTT FIPRGNQPCPCVSPISVPTSRIONVAFLLSVLPALPPTPNAMGLSPASGMP SOGLAKNTKSPQVNDNDIRTKNAVIRDLGKKTFFSDVRPNQKQYKLSSEQRLLM ETEFLERTPVDESDEIQHDEIKGCTAPIFDKRLKHFRTVEGSPVFTCKIVGIP VPVYWFQKDKOISKRNECHKRRGDEGTCSLHIESTTDDGNTYIMAAHQGRSC SGHLVOSLPISRLTSAGQSHRGSRVQERKEPLQERFRPHFLQAFQDVAHGER LCRDCKVSGLPPELUTLWLNQPVLPDASHMLVRETGVHSLLDLPDQDAGTYKC IATNKQNSFSLSELSVAKEVKAPVILEKQNGCVPEGHVRLCECRVIGMPVPVFK WKDNETIPTCTREISMHQDTTGACLLIQPAKKSADAGWTLSAKNEAGIVSCTARLD IVAQMHQIPPPMSVRPSGRVGSLSLTKGLDIFSAFSSMESTWVYSCSSRSVVEDEL "
misc_feature	1277..1531
	/note="Region: Ig-I immunoglobulin domain"
misc_feature	1790..2044
	/note="Region: Ig-I immunoglobulin domain"
misc_feature	3299..3565
	/note="Region: Ig-I immunoglobulin domain"
misc_feature	3722..3946
	/note="Region: Ig-I immunoglobulin domain"
misc_feature	4001..4246
	/note="Region: Ig-I immunoglobulin domain"
BASE COUNT	1623 a 1437 c 1294 g 1353 t
ORIGIN	
Query Match	88.4%; Score 3504; DB 9; Length 5707;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 3954; Conservative	0; Mismatches 9; Indels 0; Gaps 0;
QY	1 ATCAAGACGACAGCATAGAGCTTCTACTTCATATCTCCAGCTTCTAAGAGAGAGCTAT 60
Db	488 ATCAAGACGACAGCATAGAGCTTCTACTTCATATCTCCAGCTTCTAAGAGAGAGCTAT 547

QY 61 TTAGCTGAACACAGACATCGGGAAACAATGAGAGAGTCGAGCGGAGCCCTCCTCCAAC 120
|||||
Db 548 TTAGCTGAACACAGACATCGGGAAACAATGAGAGAGTCGAGCGGAGCCCTCCTCCAAC 607
QY 121 CCTTGGCAATTTGGCAGTCCCTTCTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 180
Db 608 CCTTGGCAATTTGGCAGTCCCTTCTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 667
QY 181 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAATATAGACGAAGTCTCAATTTGGCAAGA 240
|||||
Db 668 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAATATAGACGAAGTCTCAATTTGGCAAGA 727
QY 241 CTGGCCATCAATTTACGACCTTTTGGAGAAGGCAGATGAACCTCAAGCTAGAAAACGACTT 300
Db 728 CTGGCCATCAATTTACGACCTTTTGGAGAAGGCAGATGAACCTCAAGCTAGAAAACGACTT 787
QY 301 TCTCCTGATCAGATGAACACATCACTCACTAATTTAAAGTTTGGAGCCTAACTTCTGCCAGGAT 360
Db 788 TCTCCTGATCAGATGAACACATCACTCACTAATTTAAAGTTTGGAGCCTAACTTCTGCCAGGAT 847
QY 361 AAGCCTCGAAGTCCCAACAGCTCTAAAGAAAGCCGCCAGGAGGCAAAAAGGCCACAGTAT 420
|||||
Db 848 AAGCCTCGAAGTCCCAACAGCTCTAAAGAAAGCCGCCAGGAGGCAAAAAGGCCACAGTAT 907
QY 421 TGTCTGAAACCCAGTCCAAAAGATATTTTAAATTAAGGCTGCCGACTTCATTTGAAGAG 480
|||||
Db 908 TGTCTGAAACCCAGTCCAAAAGATATTTTAAATTAAGGCTGCCGACTTCATTTGAAGAG 967
QY 481 CTATCCTCCCTTTTCAATTCACAGCTCCAAAGATTTAGACCTCGTCCGCTGCAAAAAC 540
Db 968 CTATCCTCCCTTTTCAATTCACAGCTCCAAAGATTTAGACCTCGTCCGCTGCAAAAAC 1027
QY 541 CACAAGAGTAAAGTCTCAAAAACAAGTTATTCAGGAAACAGCTCCAGTTTCTCA 600
Db 1028 CACAAGAGTAAAGTCTCAAAAACAAGTTATTCAGGAAACAGCTCCAGTTTCTCA 1087
QY 601 GATCTGCAAGAAAGCAGAAAGATCTCTGTGTCCCATCCCTATCCCTCGGGATACCAGG 660
Db 1088 GATCTGCAAGAAAGCAGAAAGATCTCTGTGTCCCATCCCTATCCCTCGGGATACCAGG 1147
QY 661 GATAATGAAGTGAATCAGGCCCTGGAAACAGCAGGAAGCCAAAGGGCTGAAGCGGAGCAG 720
Db 1148 GATAATGAAGTGAATCAGGCCCTGGAAACAGCAGGAAGCCAAAGGGCTGAAGCGGAGCAG 1207
QY 721 GCTGCCAGTGAGCGGCTGTGTGAGACACTACACAGGGTCTTCGCCCTTCATCTCTGTAC 780
Db 1208 GCTGCCAGTGAGCGGCTGTGTGAGACACTACACAGGGTCTTCGCCCTTCATCTCTGTAC 1267
QY 781 TATGAAGAACCTCTGCGGCAACCTCCCGGTTCACTCAAAAAGTTACGGAGCAGAGAAGTT 840
Db 1268 TATGAAGAACCTCTGCGGCAACCTCCCGGTTCACTCAAAAAGTTACGGAGCAGAGAAGTT 1327
QY 841 CCAGAAGGAACCTGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACCCTCAAGTA 900
Db 1328 CCAGAAGGAACCTGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACCCTCAAGTA 1387
QY 901 AGGTGTACTGTGAAGCAGGAGCTTGAATTTCCCAAGATATTCACATCGTCCAGGCA 960
Db 1388 AGGTGTACTGTGAAGCAGGAGCTTGAATTTCCCAAGATATTCACATCGTCCAGGCA 1447
QY 961 GGAATCTGCATCTACTGACCAATTCGGGAAGCCTTTGAAGAGGACACAGSACGCTATTCC 1020
Db 1448 GGAATCTGCATCTACTGACCAATTCGGGAAGCCTTTGAAGAGGACACAGSACGCTATTCC 1507
QY 1021 TGCCTTTGCTTTAAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAGGG 1080
Db 1508 TGCCTTTGCTTTAAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAGGG 1567
QY 1081 GTTTCCTTCTTCTGACTCAGAGGCGACCTTAACAAGGAGAGATGAATCGAATCCAGAAG 1140
Db 1568 GTTTCCTTCTTCTGACTCAGAGGCGACCTTAACAAGGAGAGATGAATCGAATCCAGAAG 1627
QY 1141 CCAAAATGAGGTGTATCTCTCCCACTACCTCTGCAGTCAATTCCTCCAGCAGTACCCCA 1200
|||||

Db 1628 CCAAAATGAGGTGTATCTCTCCCACTACCTCTGCAGTCAATTCCTCCAGCAGTACCCCAA 1687
QY 1201 GCCCAGCATTTTGGTGGCCCAACCTCGTGTGGCAACCATCCAGCAGTGTGAGAGCCCAACC 1260
Db 1688 GCCCAGCATTTTGGTGGCCCAACCTCGTGTGGCAACCATCCAGCAGTGTGAGAGCCCAACC 1747
QY 1261 AATTACTTTCAGGGATTGGATGGAAACCTATCATTTGCAGCTCCTGTGTTCACAAAGATG 1320
Db 1748 AATTACTTTCAGGGATTGGATGGAAACCTATCATTTGCAGCTCCTGTGTTCACAAAGATG 1807
QY 1321 CTACAAAATTTGTACGCTTCTGAGGTCAGTGGTGTCTTTGAATGCGAGGTAAAGGA 1380
Db 1808 CTACAAAATTTGTACGCTTCTGAGGTCAGTGGTGTCTTTGAATGCGAGGTAAAGGA 1867
QY 1381 GCTCCATCTCTTAAGGTTGAGTGGTATAGAGAAGGACATTTAATAGAAGATTTCCAGAT 1440
Db 1868 GCTCCATCTCTTAAGGTTGAGTGGTATAGAGAAGGACATTTAATAGAAGATTTCCAGAT 1927
QY 1441 TTTAGGATTTTACAGAAAAAACCTCGATCCATGGCAGAGCCAGAGGAGATTTTCACCTTG 1500
Db 1928 TTTAGGATTTTACAGAAAAAACCTCGATCCATGGCAGAGCCAGAGGAGATTTTCACCTTG 1987
QY 1501 GTCATTTGCTGAGGTGTTTGCAGAAAGATTTCTGGGTGCTTCACATGTACTGCAAGCAACAAA 1560
Db 1988 GTCATTTGCTGAGGTGTTTGCAGAAAGATTTCTGGGTGCTTCACATGTACTGCAAGCAACAAA 2047
QY 1561 TAGGGCAGAGTGTCAAGCATTTGCACAGCTGCACAGTGCAGAGGAAATGAGGACCTCAGCAAC 1620
Db 2048 TAGGGCAGAGTGTCAAGCATTTGCACAGCTGCACAGTGCAGAGGAAATGAGGACCTCAGCAAC 2107
QY 1621 AAGGGTCTCTTCACTCAGCCAACTCYACCAACCACTGCGAGCTATTGAGGCCACAGCCC 1680
Db 2108 AAGGGTCTCTTCACTCAGCCAACTCTACCAACCACTGCGAGCTATTGAGGCCACAGCCC 2167
QY 1681 TCCCCACCCCACTCAGAGCTTCCATCTGTGGAAACACCCCAACCCCAACCCCAACCCAGGGG 1740
Db 2168 TCCCCACCCCACTCAGAGCTTCCATCTGTGGAAACACCCCAACCCCAACCCCAACCCAGGGG 2227
QY 1741 GTTCTGCTGAACACCAATGAGCCCGCTGCAGCTCCAGATTTGGGCTTCTGTGTGCACTTC 1800
Db 2228 GTTCTGCTGAACACCAATGAGCCCGCTGCAGCTCCAGATTTGGGCTTCTGTGTGCACTTC 2287
QY 1801 AACTCGCTGAAGTACAAAGGAAGTGAAGCATCTCCGAGGCTGTGTGTGTGACCAACC 1860
Db 2288 AACTCGCTGAAGTACAAAGGAAGTGAAGCATCTCCGAGGCTGTGTGTGTGACCAACC 2347
QY 1861 AGACAGCAGGCGCCGATTTTSCAGGAGAGTTCAACGGACAGGTTCAACGGAACCAACCAACCCCA 1920
Db 2348 AGACAGCAGGCGCCGATTTTSCAGGAGAGTTTCAACGGACAGGTTCAACGGAACCAACCAACCCCA 2407
QY 1921 GAGCCTTTCTTCCCGTGAAGAGCCCGCTCCAGTCTTGGCCAAACCCCAACCTTGATTCC 1980
Db 2408 GAGCCTTTCTTCCCGTGAAGAGCCCGCTCCAGTCTTGGCCAAACCCCAACCTTGATTCC 2467
QY 1981 ACTCAGTTCAACAGCTTCAATAACCAAGTCTTACTGGAACAACCAACCAATTCGCAAAACCCA 2040
Db 2468 ACTCAGTTCAACAGCTTCAATAACCAAGTCTTACTGGAACAACCAACCAATTCGCAAAACCCA 2527
QY 2041 CCTCCTTCATCTCCTAAGGAGTTTCTTTCARCATGACTGTTTGAACCTCAATGCTTCC 2100
Db 2528 CCTCCTTCATCTCCTAAGGAGTTTCTTTCARCATGACTGTTTGAACCTCAATGCTTCC 2587
QY 2101 CCAGCGGTGACACATCCCAATGAGCAGGTGAAGGCTCTTTCATCAGAGAGCTTCAGCTTG 2160
Db 2588 CCAGCGGTGACACATCCCAATGAGCAGGTGAAGGCTCTTTCATCAGAGAGCTTCAGCTTG 2647
QY 2161 GCCCGGCGGAAGTATTTCTTCCCTCCACGAACACACCCAGCAACTGTGGCCCTTCC 2220
Db 2648 GCCCGGCGGAAGTATTTCTTCCCTCCACGAACACACCCAGCAACTGTGGCCCTTCC 2707
QY 2221 AGCTCTCCGGTGTCTACTTTGAGCAGCACTCTCAAACCTATTTCAGAGGACAGTGAACAAA 2280
|||||

Dbb 2708 AGCTCTCCGGTGTTCACCTTTGAGCAGACCTCTCTAAACTATTCAGAGGACAGTGAACAAA 2767
Qy 2281 GAAAGCCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAAAATCTCCAGGAGGCTTTCCATC 2340
Dbb 2768 GAAAGCCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAAAATCTCCAGGAGGCTTTCCATC 2827
Qy 2341 CAAATGAGCCACTCCACCAGGCCCCAACAGAACCCACACACCAACCAATTCACATTTTCC 2400
Dbb 2828 CAAATGAGCCACTCCACCAGGCCCCAACAGAACCCACACACCAACCAATTCACATTTTCC 2887
Qy 2401 ATCCCCAGGGAACACAGTTTCAGCCCCGCTGTGTGCCCAATTCCTGTCTCTCTTACC 2460
Dbb 2888 ATCCCCAGGGAACACAGTTTCAGCCCCGCTGTGTGCCCAATTCCTGTCTCTCTTACC 2947
Qy 2461 AGCCGGATTGAGAACCCAGTGGCTTCTCAGCTCTCTGTCTGCTCTCTCTCTCTCTCTCT 2520
Dbb 2948 AGCCGGATTGAGAACCCAGTGGCTTCTCAGCTCTCTGTCTGCTCTCTCTCTCTCTCTCT 3007
Qy 2521 CCACCCCAAAATGCCATGGGCTGCGCTAGAGTGCACCATCCATGCCATCCAGGGATTA 2580
Dbb 3008 CCACCCCAAAATGCCATGGGCTGCGCTAGAGTGCACCATCCATGCCATCCAGGGATTA 3067
Qy 2581 GCGAAGAAAATACAAAGTCTCTCAACAGTGAATGATGATACATTCGTGAAACTAAG 2640
Dbb 3068 GCGAAGAAAATACAAAGTCTCTCAACAGTGAATGATGATACATTCGTGAAACTAAG 3127
Qy 2641 AACCCAGTGTTCAGAGCTTTGGGGAATAAATTAACCTTTCAGTGTATGTCAGACCAAAACAG 2700
Dbb 3128 AACCCAGTGTTCAGAGCTTTGGGGAATAAATTAACCTTTCAGTGTATGTCAGACCAAAACAG 3187
Qy 2701 CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAAATAGAGTTTCGGCTTG 2760
Dbb 3188 CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAAATAGAGTTTCGGCTTG 3247
Qy 2761 GAAGCTACTCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2820
Dbb 3248 GAAGCTACTCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3307
Qy 2821 AAGTGTATTCCTCCATCTTTGACAAGAGACTCAAGCAGTTCGGGGTTCACAGAGGCTCT 2880
Dbb 3308 AAGTGTATTCCTCCATCTTTGACAAGAGACTCAAGCAGTTCGGGGTTCACAGAGGCTCT 3367
Qy 2881 CCAGTACATTCACCTGCAAAATTTGGGATACCTGTTCCAAAGGTTTACTGGTTCAAA 2940
Dbb 3368 CCAGTACATTCACCTGCAAAATTTGGGATACCTGTTCCAAAGGTTTACTGGTTCAAA 3427
Qy 2941 GATGGGAAGCAGATTTCTAAGAGAAATGAGCAGTGCACCAATGAGCGCAGAGGATGGG 3000
Dbb 3428 GATGGGAAGCAGATTTCTAAGAGAAATGAGCAGTGCACCAATGAGCGCAGAGGATGGG 3487
Qy 3001 ACATGCTCTCTGCACATTTGAATCCACTACCACTGATGACGATGGCAACTACACCATCATG 3060
Dbb 3488 ACATGCTCTCTGCACATTTGAATCCACTACCACTGATGACGATGGCAACTACACCATCATG 3547
Qy 3061 GCAGCCAAACCCAGGGAGAAATCAGCTGTCTGGCCACATTTGATGGTACAAAGTTTGCCC 3120
Dbb 3548 GCAGCCAAACCCAGGGAGAAATCAGCTGTCTGGCCACATTTGATGGTACAAAGTTTGCCC 3607
Qy 3121 ATTGCGAGTGGCTTAACCTCTGCTGGTCACTCTCACAGGGAAGATCCCGAGTGCAGAA 3180
Dbb 3608 ATTGCGAGTGGCTTAACCTCTGCTGGTCACTCTCACAGGGAAGATCCCGAGTGCAGAA 3667
Qy 3181 AGAGACAAGAGCCCCCTACAGGAACGCTTTTCCGACACATTTCTTCGAGGCTCTCTGGG 3240
Dbb 3668 AGAGACAAGAGCCCCCTACAGGAACGCTTTTCCGACACATTTCTTCGAGGCTCTCTGGG 3727
Qy 3241 GATATGTAGTCAATGAGGGCGCTCTGTGCGTGGACGTGTAAGGTGAGTGGTTTACC 3300
Dbb 3728 GATATGTAGTCAATGAGGGCGCTCTGTGCGTGGACGTGTAAGGTGAGTGGTTTACC 3787
Qy 3301 CCCCCGAGCTGACATGGCTACTCAATGGCCAACTCTGTCTACAGATGCTCCCAACAG 3360
Dbb 3788 CCCCCGAGCTGACATGGCTACTCAATGGCCAACTCTGTCTACAGATGCTCCCAACAG 3847

Qy 3361 ATGCTGTGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCGGAC 3420
Dbb 3848 ATGCTGTGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCGGAC 3907
Qy 3421 GCAGGGACCTTAAAGTGCATCGCTACACAAACCCGGGAGAAATTTCTTTAGTCTGGAG 3480
Dbb 3908 GCAGGGACCTTAAAGTGCATCGCTACACAAACCCGGGAGAAATTTCTTTAGTCTGGAG 3967
Qy 3481 CTCTCTGTAGTACCCAAAGAGGTGAAGAAAGACACCTGTGATCTCTGGAGAAACTACAGAAC 3540
Dbb 3968 CTCTCTGTAGTACCCAAAGAGGTGAAGAAAGACACCTGTGATCTCTGGAGAAACTACAGAAC 4027
Qy 3541 TGGCGGTGTTCCCAAGGCCACCCCGTGGAGTGGAGTGGCGGTGATAGGCATGCCCCCA 3600
Dbb 4028 TGGCGGTGTTCCCAAGGCCACCCCGTGGAGTGGAGTGGCGGTGATAGGCATGCCCCCA 4087
Qy 3601 CTTGTGTTCTACTGGAGAAAGACAATGAGACCATCCCTTTCACACAGAGAGAGATCAGT 3660
Dbb 4088 CTTGTGTTCTACTGGAAGAAAGACAATGAGACCATCCCTTTCACACAGAGAGAGATCAGT 4147
Qy 3661 ATGCACAGGACACACAGGGTATGCTGCTCTCATTCAGCCAGCCAAAGAAATCAGAC 3720
Dbb 4148 ATGCACAGGACACACAGGGTATGCTGCTCTCATTCAGCCAGCCAAAGAAATCAGAC 4207
Qy 3721 GCTGGATGTTACACGTTGTCAGCAAGAATGAAGCCGGCATCGTGTGCTGCTGCTGCTGCTG 3780
Dbb 4208 GCTGGATGTTACACGTTGTCAGCAAGAATGAAGCCGGCATCGTGTGCTGCTGCTGCTGCTG 4267
Qy 3781 CTGGATATATACGCTCAGTGGCCACCATCAGATCCACCGCCCATGCTGTGTCGGCCCATG 3840
Dbb 4268 CTGGATATATACGCTCAGTGGCCACCATCAGATCCACCGCCCATGCTGTGTCGGCCCATG 4327
Qy 3841 GGCAGTGCCTACGATCTCTCAGCAGTAAAGAGCTTGACATATTTTCTGCTTTTCTCTCC 3900
Dbb 4328 GGCAGTGCCTACGATCTCTCAGCAGTAAAGAGCTTGACATATTTTCTGCTTTTCTCTCC 4387
Qy 3901 ATGGAAGACGATGTTGATTCATGCTCTTCTCGGAGTGTAGTGGAGAGTGTGTAACCTT 3960
Dbb 4388 ATGGAAGACGATGTTGATTCATGCTCTTCTCGGAGTGTAGTGGAGAGTGTGTAACCTT 4447
Qy 3961 TAA 3963
Dbb 4448 TAA 4450

RESULT 2
HSM805275
LOCUS
DEFINITION
Homo sapiens mRNA; cDNA DKF2p451e012 (from clone DKF2p451e012).
ACCESSION
AL834247
VERSION
AL834247.1 GI:21739789
KEYWORDS
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 5804)
AUTHORS
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE
Direct Submission
JOURNAL
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKF2p451e012) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
Information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.
FEATURES
source
1..5804
/organism="Homo sapiens"

```
/db_xref="taxon:9606"
/chromosome="10"
/clone="DKF2p451E012"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..5804
/gene="DKF2p451E012"
<..4227
/gene="DKF2p451E012"
/notes="similarity to actin-associated protein palladin
(Mus musculus)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD38923.1"
/db_xref="GI:21739790"
/translation="YGCEITTEGGADKCSGGDRLLSFLQPSLKLRLTPHSLMTAT
LYFGREPEPPSVLSSKSPWIIIAVECLDWTSSGSKKESQDDDSIEASTISQ
LLREYLAETRRGNRRSRAPNCPHFSGPAGAEQGGDDLPDLAFLSOEEL
DESVLARLAINVDPLEKADETOAKRLSPDOMKHSPLNSFFPNFCODNPRSPKSKE
SQEAKRPOYCSOTSKKVFPLNAADFIEELSLKSHSKRIRPRACKNHRKSLKESQ
NKVMNENSSFSDLSREBRKSSVPIPADTRDNEVNHAEQOEAKRRAEQAASEAA
GGDTTPGSSPDLXYEEPLGQPPRTFKLRSREVPEGTRVQDLCIVGTPPOVRMYC
EGKELENSDIIHVQAGNLHSLTIAEAFEDTGRYSCFASNIYIGDSTSAEIVIEGV
SSDSEGDPNKEEMNRIOKPNVSSPTTSAVPPAIPAQHLVAOPRVATIOQCSP
NYLOGLDGKPIIAAPVFTKMLONLSASEGLVFEKRVKGAQSPKVEWYREGTLIEDS
PFRILQKPRSMAREPEICTLIVAEVFAEDSGCTCTASNKYGTVSSIAQLHVRGNE
DUSNGLSHNSNTNLAIEPOPSPHSEPPSPVOPPAKPLEGVLVNHNEPRSSRI
LURVHNPDEDDKSSASEAGVTTTRTPDSFQERENGQATKTPSPSPVKEPPV
LAKPLDSTLOALHNLVLEQHLQNPSPSPKFPFMTVLNSNPAVATSSKQV
KAPSSOTFLARPKYFFPNTNTAATVAPSSPVPVTLSTPOTIORTVSKESLLVSH
SVQTKSPGLSTONELPPGTEPPPTFSIPSGNOFQPCVSPIPVSPSTRIONP
VAFSLVLSLPAIPTNAMGLPRSNAPMPSOGLAKKNTKSPQVNDNIRETKNAVI
ROLGKKTISDVRPNQOEYKISSEFORLWNEIEFRLETPVDESDDEIHOHDEIPTGK
IAPIDRLKHFRFTEGSPVTFCTKIVGPVKVYWFKDQKQISKRNERCKMRREGDG
TCSLHIESITDDGNTYIMAAINPOGRIKSCGHLMVQSLPIRSRLTSAGQSHRGRSV
QERDKPELREFFRHLLOAPDMVAHEGRCLDCKVSGLPPELTLWLNQOLPDP
APKMLVRETVGHSLLDPLTORDAGTYKCIATNKTQNSFSLELSVAKEVKKAPVI
LEKLQNGVPEGHVPLECRVIGMPPPVYWKNDNETIPTRERISMHQDITGYACLL
IQPAKSDAGWITLSKNAEGIVSCTARLDIYAQMHHQIPPPMSVVRPSGRYSGLTSK
GDDIFSFSMSMETMYVSCSRSVSEDEL"
polyA_signal
/gene="DKF2p451E012"
5783
polyA_site
/gene="DKF2p451E012"
BASE COUNT 1650 a 1460 c 1322 g 1372 t
ORIGIN
```

```
Query Match 87.1%; Score 3453; DB 9; Length 5804;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3953; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 ATCAAGACGACGATAGAGCTTCTTCTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 60
|
|
|
Db 265 ATCAAGACGACGATAGAGCTTCTTCTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 324
|
|
|
QY 61 TTAGCTGAACACGACATCGGGGAACAATGAGAGAGTGCAGGGAGCCCTCTCTCAAC 120
|
|
|
Db 325 TTAGCTGAACACGACATCGGGGAACAATGAGAGAGTGCAGGGAGCCCTCTCTCAAC 384
|
|
|
QY 121 CCTTGGCCATTTGCGAGCTCTCTTGGGGCGCTGAAGGAGCGAGGCCAAGATGACCTT 180
|
|
|
Db 385 CCTTGGCCATTTGCGAGCTCTCTTGGGGCGCTGAAGGAGCGAGGCCAAGATGACCTT 444
|
|
|
QY 181 CCAGATCTTTTTCAGCTTTCTGAGCCAAAGAAGATTTAGACGAAAAGTGTCAATTTGGCAAGA 240
|
|
|
Db 445 CCAGATCTTTTTCAGCTTTCTGAGCCAAAGAAGATTTAGACGAAAAGTGTCAATTTGGCAAGA 504
|
|
|
QY 241 CTGCCCATCAATTACGACCTTTTGGAGAGGCGAGATGAAACTCAAGCTAGAAACGACTT 300
|
|
|
Db 505 CTGCCCATCAATTACGACCTTTTGGAGAGGCGAGATGAAACTCAAGCTAGAAACGACTT 564
|
|
|
```

```
QY 301 TCTCTCTGATCAGATGAACACTCACCTAAATTTAAGTTTTCGACCTAACTTCTGCCAGGAT 360
|
|
|
Db 565 TCTCTCTGATCAGATGAACACTCACCTAAATTTAAGTTTTCGACCTAACTTCTGCCAGGAT 624
|
|
|
QY 361 AACCTCTGAAGTCCCACCAGCTCTAAAGAAAGCCCCAGGAGGCAAAAAGGCCACAGTAT 420
|
|
|
Db 625 AACCTCTGAAGTCCCACCAGCTCTAAAGAAAGCCCCAGGAGGCAAAAAGGCCACAGTAT 684
|
|
|
QY 421 TGTCTCTGAACACCAGTCCCAAAAAGTATTTTAAATAAGGCTCCCGACTTCATTGAAGAG 480
|
|
|
Db 685 TGTCTCTGAACACCAGTCCCAAAAAGTATTTTAAATAAGGCTCCCGACTTCATTGAAGAG 744
|
|
|
QY 481 CTATCTCTCCCTTTTCAAAATCCCACAGCTCCAAAAGGATTAGACCTCGTGCCTGCACAAAAC 540
|
|
|
Db 745 CTATCTCTCCCTTTTCAAAATCCCACAGCTCCAAAAGGATTAGACCTCGTGCCTGCACAAAAC 804
|
|
|
QY 541 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTATGACGAGAAACAGCTCCAGTTTCTCA 600
|
|
|
Db 805 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTATGACGAGAAACAGCTCCAGTTTCTCA 864
|
|
|
QY 601 GATCTGTCAAGACGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTCGGGATACGAG 660
|
|
|
Db 865 GATCTGTCAAGACGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTCGGGATACGAG 924
|
|
|
QY 661 GATAATGAAGTGAATCACGCCCTTGGAAACAGCAGAGAGCCCAAGAGGCGTGAAGCGGAGCAG 720
|
|
|
Db 925 GATAATGAAGTGAATCACGCCCTTGGAAACAGCAGAGAGCCCAAGAGGCGTGAAGCGGAGCAG 984
|
|
|
QY 721 GCTGCCAGTGAAGCGGCTGGTGAGACACTACACAGGGTCTTCCCTTCTCATCTCTGTATC 780
|
|
|
Db 985 GCTGCCAGTGAAGCGGCTGGTGAGACACTACACAGGGTCTTCCCTTCTCATCTCTGTATC 1044
|
|
|
QY 781 TATGAAGAACTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTACGAGACGAGAAAGTT 840
|
|
|
Db 1045 TATGAAGAACTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTACGAGACGAGAAAGTT 1104
|
|
|
QY 841 CCAGAAGAACTCGAGTACAGTTGGATTGATAGTGGTAGGAAATCCACCACTCAAGTA 900
|
|
|
Db 1105 CCAGAAGAACTCGAGTACAGTTGGATTGATAGTGGTAGGAAATCCACCACTCAAGTA 1164
|
|
|
QY 901 AGTGGTACTGTGAAGCAAGAGAGCTTTGAAAATTTCCCAAGATATTCATCTGTCAGGCA 960
|
|
|
Db 1165 AGTGGTACTGTGAAGCAAGAGAGCTTTGAAAATTTCCCAAGATATTCATCTGTCAGGCA 1224
|
|
|
QY 961 GGAATCTGCACCTACCTGACCATTTGGGAAGCCCTTTGAAGAGGACAGACGAGCTATTC 1020
|
|
|
Db 1225 GGAATCTGCACCTACCTGACCATTTGGGAAGCCCTTTGAAGAGGACAGACGAGCTATTC 1284
|
|
|
QY 1021 TGCTTTGCTTCTTAACATCTATGGGACAGATTGAGACTTCTGCTGAGATTTATATAGAAGGG 1080
|
|
|
Db 1285 TGCTTTGCTTCTTAACATCTATGGGACAGATTGAGACTTCTGCTGAGATTTATATAGAAGGG 1344
|
|
|
QY 1081 GTTCTTCTTCTTCTGACTCAGAAGCGACCCCTAACAAAGGAAGAGATGAATCGAATCCAGAAG 1140
|
|
|
Db 1345 GTTCTTCTTCTTCTGACTCAGAAGCGACCCCTAACAAAGGAAGAGATGAATCGAATCCAGAAG 1404
|
|
|
QY 1141 CCAATGAGTGTCTATCTCTCCACTACCTCTGCACTCATTCCTGCACTGATTCCTCAGCAGTACCCCAA 1200
|
|
|
Db 1405 CCAATGAGTGTCTATCTCTCCACTACCTCTGCACTCATTCCTCAGCAGTACCCCAA 1464
|
|
|
QY 1201 GCCCAGCATTTGGTGGCCCAACCTCTGTTGGCAACCATCCAGCAGTGTCCAGAGCCCCACC 1260
|
|
|
Db 1465 GCCCAGCATTTGGTGGCCCAACCTCTGTTGGCAACCATCCAGCAGTGTCCAGAGCCCCACC 1524
|
|
|
QY 1261 AATTACTTGCAGGATTTGGATGGAACACCTATCATTTGACGCTCCTGTTGTTTACAAGATG 1320
|
|
|
Db 1525 AATTACTTGCAGGATTTGGATGGAACACCTATCATTTGACGCTCCTGTTGTTTACAAGATG 1584
|
|
|
QY 1321 CTACAAATTTGTCAGCTTCTGAGGGTCACTGGTGTCTTTGTAATGCAGATGAAAAGGA 1380
|
|
|
Db 1585 CTACAAATTTGTCAGCTTCTGAGGGTCACTGGTGTCTTTGTAATGCAGATGAAAAGGA 1644
|
|
|
QY 1381 GCTCCATCTCTTAGGTTGAGTGGTATAGAGAAGGGACTTTAATAGAAGATTCTCCAGAT 1440
|
|
|
```

[illegible]

Db	2725	AGCGGATTCAGAACCCAGTGGCTTTCCTCAGCTCTGTTCTCTGCCCTCTCTCCTCCCTGCCCATC	2784
Qy	2521	CCACCCAAATGCCATCTGGRGCTGCCTAGAAAGTGCACCATCCATGCCATCCACAGGATTA	2580
Db	2785	CCACCCAAATGCCATCTGGGCTGCCTAGAAAGTGCACCATCCATGCCATCCACAGGATTA	2844
Qy	2581	CGAAGAAAAATACAAAGTCTCCTCAACAGTGAATGATGATAACATTCGTGGAACCTAAG	2640
Db	2845	CGAAGAAAAATACAAAGTCTCCTCAACAGTGAATGATGATAACATTCGTGGAACCTAAG	2904
Qy	2641	AACSCAGTGATTCGAGACTTGGGGAANAATAACTTTCAGTGATGTCAGACCAACACAG	2700
Db	2905	AACSCAGTGATTCGAGACTTGGGGAANAATAACTTTCAGTGATGTCAGACCAACACAG	2964
Qy	2701	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGTTTCGCTTG	2760
Db	2965	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGTTTCGCTTG	3024
Qy	2761	GAAGCTACTCCTGTTGATGAATCAGATGATGAATTAACATGATGAGATCCCCACGGGC	2820
Db	3025	GAAGCTACTCCTGTTGATGAATCAGATGATGAATTAACATGATGAGATCCCCACGGGC	3084
Qy	2821	AAGTGTATTGCTCCCATCTTTTGACAGAGACTCAAGCACTTCCGGGTGACAGAAAGGCTCT	2880
Db	3085	AAGTGTATTGCTCCCATCTTTTGACAGAGACTCAAGCACTTCCGGGTGACAGAAAGGCTCT	3144
Qy	2881	CCAGTCAATTCACCTGCAAAATTTGTTGGATACCTGTTTCAAAGGTTTACTGGTTCAAA	2940
Db	3145	CCAGTCAATTCACCTGCAAAATTTGTTGGATACCTGTTTCAAAGGTTTACTGGTTCAAA	3204
Qy	2941	GATGGGAAGCAGATTTTCTAAGAGAAATAGCACTGCAAAATGAGCGAGAGAGAGATGGG	3000
Db	3205	GATGGGAAGCAGATTTTCTAAGAGAAATAGCACTGCAAAATGAGCGAGAGAGAGATGGG	3264
Qy	3001	ACATGCTCTGTGCACATGTAATCCACTACAGTGATGAGATGGCACTACACATCATG	3060
Db	3265	ACATGCTCTGTGCACATGTAATCCACTACAGTGATGAGATGGCACTACACATCATG	3324
Qy	3061	CGAGCAACCCAGGAGGAATCAGCTGTCTGGCCACTGATGGTACAAAGTTTGCC	3120
Db	3325	CGAGCAACCCAGGAGGAATCAGCTGTCTGGCCACTGATGGTACAAAGTTTGCC	3384
Qy	3121	ATTCCGAGTCGGCTTAACCTCTGCTGGTCTAGCTCTCACAGGGGAAGATCCCGAGTGCAGAA	3180
Db	3385	ATTCCGAGTCGGCTTAACCTCTGCTGGTCTAGCTCTCACAGGGGAAGATCCCGAGTGCAGAA	3444
Qy	3181	AGAGACAAAGAGCCCTTACAGGAACCGTTTTTCCGACACATTTCTCGAGGCTCTGGG	3240
Db	3445	AGAGACAAAGAGCCCTTACAGGAACCGTTTTTCCGACACATTTCTCGAGGCTCTGGG	3504
Qy	3241	GATATGTAGCTCATGAGGGCGCCTCTGTGGCTGGACTGTAAAGTGAGTGGTTTACCG	3300
Db	3505	GATATGTAGCTCATGAGGGCGCCTCTGTGGCTGGACTGTAAAGTGAGTGGTTTACCG	3564
Qy	3301	CCCCGGAGCTGACATGGCTTACTCAATGGCCAACTGTGCTACCAAGATGCCTCCCAACAG	3360
Db	3565	CCCCGGAGCTGACATGGCTTACTCAATGGCCAACTGTGCTACCAAGATGCCTCCCAACAG	3624
Qy	3361	ATGCTGTCAGGGAGACCGGAGTCCACTCTGCTCATTTGACCCACTCACTCAGCGGAC	3420
Db	3625	ATGCTGTCAGGGAGACCGGAGTCCACTCTGCTCATTTGACCCACTCACTCAGCGGAC	3684
Qy	3421	CGAGGGACCTTAAAGTGATCGCTACCAACAAAACGGGCGAGAAATTTCTTTTAGTCTGGAG	3480
Db	3685	CGAGGGACCTTAAAGTGATCGCTACCAACAAAACGGGCGAGAAATTTCTTTTAGTCTGGAG	3744
Qy	3481	CTCTCTGTAGTACCAAGAGGTGAAGAAGCACTGTGATCTCGAGAAACTACAGAAC	3540
Db	3745	CTCTCTGTAGTACCAAGAGGTGAAGAAGCACTGTGATCTCGAGAAACTACAGAAC	3804
Qy	3541	TGCGGTGTTCCGAAGGCCACCCGCTGAGACTGGATGCGCGGTGATAGGATGCGCCCA	3600
Db	3805	TGCGGTGTTCCGAAGGCCACCCGCTGAGACTGGATGCGCGGTGATAGGATGCGCCCA	3864

```
QY 3601 CCTGTGTTCTACTCGAAGAAAGACAATGAGACCATCCCTTGCACACAGAGAGGATCAGT 3660
|||||
Db 3865 CCTGTGTTCTACTCGAAGAAAGACAATGAGACCATCCCTTGCACACAGAGAGGATCAGT 3924
|||||
QY 3661 ATGCACAGAGACACAACAGGGTATGCTGCTTCTCATTCAGCCAGCCAAAGAAATCAGAC 3720
|||||
Db 3925 ATGCACAGAGACACAACAGGGTATGCTGCTTCTCATTCAGCCAGCCAAAGAAATCAGAC 3984
|||||
QY 3721 GCTGGATGTTACACGTTGTTCAGCCAAAGATGAAGCCGGCATCTGTCGTGCATCGCCAGG 3780
|||||
Db 3985 GCTGGATGTTACACGTTGTTCAGCCAAAGATGAAGCCGGCATCTGTCGTGCATCGCCAGG 4044
|||||
QY 3781 CTGGATATATACGCTCAGTGGCCACCATCAGATCCCAACCGCCCATGTCTGTCGCCGCCAGT 3840
|||||
Db 4045 CTGGATATATACGCTCAGTGGCCACCATCAGATCCCAACCGCCCATGTCTGTCGCCGCCAGT 4104
|||||
QY 3841 GGCAGTCGCTACGGATCTCTCAGCAGTAAGGACTTGACATATTTTCTGCTTTTCTCTCC 3900
|||||
Db 4105 GGCAGTCGCTACGGATCTCTCAGCAGTAAGGACTTGACATATTTTCTGCTTTTCTCTCC 4164
|||||
QY 3901 ATGGAAGCAGCATGTTGTTATTCATGCTCTCTCGGAGTCTAGTGGAGATGATGAACTT 3960
|||||
Db 4165 ATGGAAGCAGCATGTTGTTATTCATGCTCTCTCGGAGTCTAGTGGAGATGATGAACTT 4224
|||||
QY 3961 TAA 3963
|||
Db 4225 TAA 4227
|||||

RESULT 3
HSM803309 5486 bp mRNA linear PRI 12-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp451M023 (from clone DKFZp451M023).
DEFINITION AL832002
ACCESSION
VERSION AL832002.1 GI:21732542
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5486)
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.
Direct Submission
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AcOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp451M023) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cbna/.
FEATURES
Location/Qualifiers
source
1..5486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp451M023"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
pH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal 5451..5456
polyA_site 5468
BASE COUNT 1558 a 1392 c 1260 g 1276 t
ORIGIN
Query Match 82.0%; Score 3249; DB 9; Length 5486;
Best Local Similarity 99.6%; Pred.No. 0;
Matches 3949; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 ATGCAAGACGACGACATAGAAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 60
|||||
```

```
Db 266 ATGCAAGACGACGACATAGAAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 325
|||||
QY 61 TTAGCTGAACACAGACATCGGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCTCAAC 120
|||||
Db 326 TTAGCTGAACACAGACATCGGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCTCAAC 385
|||||
QY 121 CCTTGGCATTTTCGGCAGTCCCTCTGGGGCCGCTGAAGGAGCGGAGCCCAAGATGACCTT 180
|||||
Db 386 CCTTGGCATTTTCGGCAGTCCCTCTGGGGCCGCTGAAGGAGCGGAGCCCAAGATGACCTT 445
|||||
QY 181 CCAGATCTTTTCAGCCCTTTTCAGCCCAAGAAGAATTAGACGAAAGTGTCAATTTGGCAAGA 240
|||||
Db 446 CCAGATCTTTTCAGCCCTTTTCAGCCCAAGAAGAATTAGACGAAAGTGTCAATTTGGCAAGA 505
|||||
QY 241 CTGGCCATCAATTTACGACCCTTTTGGAGAGGCGAGATCAAACTCAAGCTAGAAACGACTT 300
|||||
Db 506 CTGGCCATCAATTTACGACCCTTTTGGAGAGGCGAGATGAAACTCAAGCTAGAAACGACTT 565
|||||
QY 301 TCTCCTGATCAGATGAAACACTCACCTTAATTTAAGTTTTTTCAGGCTAACTTCTGCGCAGGAT 360
|||||
Db 566 TCTCCTGATCAGATGAAACACTCACCTTAATTTAAGTTTTTTCAGGCTAACTTCTGCGCAGGAT 625
|||||
QY 361 AACCCCTCGAAGTCCACACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT 420
|||||
Db 626 AACCCCTCGAAGTCCACACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT 685
|||||
QY 421 TGTTCCTGAAACCCAGTCCCAAAAGATTTTAAATAAGGCTGCCGACTCANTTGAAGAG 480
|||||
Db 686 TGTTCCTGAAACCCAGTCCCAAAAGATTTTAAATAAGGCTGCCGACTCANTTGAAGAG 745
|||||
QY 481 CTATCTCCTCTTCAAAATCCACAGCTCCAAAGGATTAGACCTCGTGCCTGCAAAAAC 540
|||||
Db 746 CTATCTCCTCTTCAAAATCCACAGCTCCCAAGGATTAGACCTCGTGCCTGCAAAAAC 805
|||||
QY 541 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTTATGCAGGAAACACAGCTCCAGTTTCTCA 600
|||||
Db 806 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTTATGCAGGAAACACAGCTCCAGTTTCTCA 865
|||||
QY 601 GATCTGTCAGAAAGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTGGGGATACCAGG 660
|||||
Db 866 GATCTGTCAGAAAGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTGGGGATACCAGG 925
|||||
QY 661 GATAAATGAAGTGAATCACGCGCTGGAAACAGCAGCAAGCGGCTGAAGCGGAGCAG 720
|||||
Db 926 GATAAATGAAGTGAATCACGCGCTGGAAACAGCAGCAAGCGGCTGAAGCGGAGCAG 985
|||||
QY 721 GCTGCCAGTGAAGCGGCTGGTGAGACACTACACAGGGTCTTCCCTCTCATCTCTGTAC 780
|||||
Db 986 GCTGCCAGTGAAGCGGCTGGTGAGACACTACACAGGGTCTTCCCTCTCATCTCTGTAC 1045
|||||
QY 781 TATGAAGAACTCTGGGCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGAAGTT 840
|||||
Db 1046 TATGAAGAACTCTGGGCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGAAGTT 1105
|||||
QY 841 CCAGAGGAACTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACCACCTCAAGTA 900
|||||
Db 1106 CCAGAGGAACTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACCACCTCAAGTA 1165
|||||
QY 901 AGTGGTACTGTGAAGCAAGAGACTTGAATAATTCCTCCAGATATTCACATCGTCCAGGCA 960
|||||
Db 1166 AGTGGTACTGTGAAGCAAGAGACTTGAATAATTCCTCCAGATATTCACATCGTCCAGGCA 1225
|||||
QY 961 GGAATCTGCACCTCACTGACCATTTGGCGAAGCCCTTGAAGAGGACAGACGACGCTATTC 1020
|||||
Db 1226 GGAATCTGCACCTCACTGACCATTTGGCGAAGCCCTTGAAGAGGACAGACGACGCTATTC 1285
|||||
QY 1021 TGCTTTGCTTCTTAACATCTATGGGACAGATTTCGACTTCTGCTGAGATTTATATAGAAGGG 1080
|||||
Db 1286 TGCTTTGCTTCTTAACATCTATGGGACAGATTTCGACTTCTGCTGAGATTTATATAGAAGGG 1345
|||||
QY 1081 GTTTCCTTCTTCTGACTCAGAGGCGACCCCTAAACAAGGAAGAGATGAATCGAATCCAGAAG 1140
|||||
```



```
QY 3301 CCCCCGGAGCTGACATGGCTACTCAATGGCCAAACCTGTGTACCATGATGCTCCCAACAG 3360
Db      |
QY 3361 ATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTAGCCGGAC 3420
Db      |
QY 3421 GCAGGACCTATAAATGTCATCGCTACCAACAAACCGGGCAGAAATCTTTAGCTGGAG 3480
Db      |
QY 3481 CTCTCTGTAGTACCCAAAGAGGTGAAGAAAGCACCTGTGTATCTCTGGAGAAACTACAGAAC 3540
Db      |
QY 3541 TGGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCGCCA 3600
Db      |
QY 3601 CCTGTGTTCTACTGGAAGAAAGACAATGAGACCATCCCTTGCACACAGAGAGAGATCAGT 3660
Db      |
QY 3661 ATGCACAGGACACAAACAGGATGCTGCTCTCTCTCATTTAGCCAGCCAGCAAGAAATCAGAC 3720
Db      |
QY 3721 GCTGGATGTACAGCTTGTGACCCAGAAATGAAGCCGGCATCTGTGTCGTCATCGCAGG 3780
Db      |
QY 3781 CTGGATATATACGCTCAGTGGCCACCATCAGATCCACCGCCCATGCTGTGCCGCCCCAGT 3840
Db      |
QY 3841 GCGAGTCGCTACGGATCTCTCACCAGTAAAGGACTTGACATATTTCTGCTTTTCTCTCC 3900
Db      |
QY 3901 ATGGAAGCAGCATGTGTATTCATGCTCTTCGAGAGTAGTGAGAGTGATGAACCT 3960
Db      |
QY 3961 TAA 3963
Db      |
QY 4046 CTGGATATATACGCTCAGTGGCCACCATCAGATCCACCGCCCATGCTGTGCCGCCCCAGT 4105
Db      |
QY 4106 GCGAGTCGCTACGGATCTCTCACCAGTAAAGGACTTGACATATTTCTGCTTTTCTCTCC 4165
Db      |
QY 4166 ATGGAAGCAGCATGTGTATTCATGCTCTTCGAGAGTAGTGAGAGTGATGAACCT 4225
Db      |
QY 4226 TAA 4228
Db      |

RESULT 4
HSM803687
LOCUS
DEFINITION Homo sapiens mRNA; cDNA DKFZp451K027 (from clone DKFZp451K027).
ACCESSION AL832379
VERSION AL832379.1 GI:21732942
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Koehler,K., Beyer,A., Mewes,H.W., Well,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp451K027) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
```

```
FEATURES
Source
1..4683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp451K027"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal
polyA_site
BASE COUNT 1376 a 1147 c 1048 g 1112 t
ORIGIN
Query Match 55.7%; Score 2602; DB 9; Length 4683;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3052; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 903 GTGGTACTGTGAAGCAAGGAGCTTGAATAATCCCCAGATATTCCACATCTGCCAGGCAGG 962
Db      |
QY 963 AATCTGCACCTACTGACCAATTCGGGAAGCCTTTGAAGAGGACACAGGACGCTATTTCCTG 1022
Db      |
QY 1023 CTTTGTCTTAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAGGGGT 1082
Db      |
QY 412 CTTTGTCTTAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAGGGGT 471
QY 1083 TTCTTCTTCTGACTCAGAAGCGCACCTAAACAGGAAGAGATGAATCGAATCCAGAAGCC 1142
Db      |
QY 472 TTCTTCTTCTGACTCAGAAGCGCACCTAAACAGGAAGAGATGAATCGAATCCAGAAGCC 531
QY 1143 AATAGAGGTGTCATCTCTCCCACTACCTCTGCACTCATTTCTCCAGCAGTACCCCAAGC 1202
Db      |
QY 532 AATAGAGGTGTCATCTCTCCCACTACCTCTGCACTCATTTCTCCAGCAGTACCCCAAGC 591
QY 1203 CCAGCATTTGGTGGCCCCAACCTCGTGTGGCAACCATCCAGCAGTGTCCAGAGCCCCACAA 1262
Db      |
QY 592 CCAGCATTTGGTGGCCCCAACCTCGTGTGGCAACCATCCAGCAGTGTCCAGAGCCCCACAA 651
QY 1263 TTACTTGCAGGGATTTGGATGGAACCTATCATTGCACTCTCTGTTTACAAAGATGCT 1322
Db      |
QY 652 TTACTTGCAGGGATTTGGATGGAACCTATCATTGCACTCTCTGTTTACAAAGATGCT 711
QY 1323 ACAAATTTTGCAGCTTCTGAGGGTCACTGCTGTTGCTTTTGAATGCAGAGTAAAGGAGC 1382
Db      |
QY 712 ACAAATTTTGCAGCTTCTGAGGGTCACTGCTGTTGCTTTTGAATGCAGAGTAAAGGAGC 771
QY 1383 TCCATCTCTTAAGTTGGAGTGGTATAGAGAGGGACCTTTATAGAGATTTCTCCAGATTT 1442
Db      |
QY 772 TCCATCTCTTAAGTTGGAGTGGTATAGAGAGGGACCTTTATAGAGATTTCTCCAGATTT 831
QY 1443 TAGGATTTTACAGAAAAAACCTCGATCCATGGCAGAGCCAGAGAGATTTGCACCTTGGT 1502
Db      |
QY 832 TAGGATTTTACAGAAAAAACCTCGATCCATGGCAGAGCCAGAGAGATTTGCACCTTGGT 891
QY 1503 CATTTGCTGAGGTGTTTTCAGAGAATTTCTGGGTGCTTCCATGTACTGCAAGCAACAATA 1562
Db      |
QY 892 CATTTGCTGAGGTGTTTTCAGAGAATTTCTGGGTGCTTCCATGTACTGCAAGCAACAATA 951
QY 1563 CGGCACAGTGTCAAGCATTCACAGCTGCAGCTGAGAGAGAAATGAGACCTCAGCAACAA 1622
Db      |
QY 952 CGGCACAGTGTCAAGCATTCACAGCTGCAGCTGAGAGAGAAATGAGACCTCAGCAACAA 1011
QY 1623 CGGTGCTCTTCACTCAGCCCAACTCYACCACCACTGCGAGCTATTGAGCCACAGCCCTC 1682
Db      |
QY 1012 CGGTGCTCTTCACTCAGCCCAACTCTACCCCACTGGCAGCTATTGAGCCACAGCCCTC 1071
QY 1683 CCCACCCCACTCAGAGCTCCATCTGTGGAAACAACCCCAACCCCAACCTCAGAGGGGT 1742
Db      |
```

Db 1072 CCCACCCCTACTAGAGCCTCCATCTGTGGAAACAACCCCCCAAAACCCAACTCGAGGGGT 1131
Qy 1743 TCTGGTGAACACACAAATGAGCCCGGTCCAGCTCCAGGATTGGGCTTCGTGTGCACACTTCAA 1802
Db 1132 TCTGGTGAACACAAATGAGCCCGGTCCAGCTCCAGGATTGGGCTTCGTGTGCACACTTCAA 1191
Qy 1803 CCTGCCCTGAAGATGACAAAGGAAGTGAAGCATCCTCCGAGGCTGGTGTGGTGACCAACAG 1862
Db 1192 CCTGCCCTGAAGATGACAAAGGAAGTGAAGCATCCTCCGAGGCTGGTGTGGTGACCAACAG 1251
Qy 1863 ACAGACAGGCCCGATTTCTTTSCAGGAGAGTTCAACGGACAGGCACAAACACCCACAGA 1922
Db 1252 ACAGACAGGCCCGATTTCTTTCCAGGAGAGTTTCAACGGACAGGCACAAACACCCACAGA 1311
Qy 1923 GCCTTTCTTCCCGTGAAGAGCCCTCCAGTTCTCGCAAAACCCAAACTTGATTCAC 1982
Db 1312 GCCTTTCTTCCCGTGAAGAGCCCTCCAGTTCTCGCAAAACCCAAACTTGATTCAC 1371
Qy 1983 TCAGTTTACAACAGCTTCATTAACCAAGTCTTACTGGAACAACACCAATTGCAAAACCCACC 2042
Db 1372 TCAGTTTACAACAGCTTCATTAACCAAGTCTTACTGGAACAACACCAATTGCAAAACCCACC 1431
Qy 2043 TCCTTCATCTTAAGAGATTTCCTTTTCARCATGACTGTTTGAACCTCAATGCTCCGCC 2102
Db 1432 TCCTTCATCTCTTAAGAGATTTCCTTTTCAGCATGACTGTTTGAACCTCAATGCTCCGCC 1491
Qy 2103 AGCGGTGACAACATCCARTAAGCAGGTGAAGGCTCTTCATCACAGACGTTTCAGCTTGGC 2162
Db 1492 AGCGGTGACAACATCCAGTAAGCAGGTGAAGGCTCTTCATCACAGACGTTTCAGCTTGGC 1551
Qy 2163 CCGGCCGAAGTATTTCCTCCCTCCAGGAACACCAACCGCAGCAACTGTGCCCCCTCCAG 2222
Db 1552 CCGGCCGAAGTATTTCCTCCCTCCAGGAACACCAACCGCAGCAACTGTGCCCCCTCCAG 1611
Qy 2223 CTCTCCGGTGTACTTTGAGCAGCACTCTCTCAAACTATTTCAGAGGACAGTGACCAAGA 2282
Db 1612 CTCTCCGGTGTACTTTGAGCAGCACTCTCTCAAACTATTTCAGAGGACAGTGACCAAGA 1671
Qy 2283 AAGCCTCTTAGTGTCACACCCCTCTGTGCAACCAAACTCTCCAGAGGGCTTTCCATCCA 2342
Db 1672 AAGCCTCTTAGTGTCACACCCCTCTGTGCAAAACCAAACTCTCCAGAGGGCTTTCCATCCA 1731
Qy 2343 AAATGAGCCACTCCACAGGCCCAACAGAACACCAACCAACCACTTCACATTTTCCAT 2402
Db 1732 AAATGAGCCACTCCACAGGCCCAACAGAACCAACCAACCACTTCACATTTTCCAT 1791
Qy 2403 CCCCAGCGGAACAGATTTTCAGCCCCCTGTGTGTCCTCCCAATTCCTGTCTCTCTACAG 2462
Db 1792 CCCCAGCGGAACAGATTTTCAGCCCCCTGTGTGTCCTCCCAATTCCTGTCTCTCTACAG 1851
Qy 2463 CCGGATTCAGAACCCAGTGGCTTTCCCTCAGCTCTGTGTCCTGCTCTCTCCCTGCCATCCC 2522
Db 1852 CCGGATTCAGAACCCAGTGGCTTTCCCTCAGCTCTGTGTCCTGCTCTCTCCCTGCCATCCC 1911
Qy 2523 ACCACAAATGCCATGGGTGCTTAGAAGTGCACCACTCCATGCCATCCCAGGGATTAGC 2582
Db 1912 ACCACAAATGCCATGGGTGCTTAGAAGTGCACCACTCCATGCCATCCCAGGGATTAGC 1971
Qy 2583 GAAGAAAAATACAAGTCTCTCAACCAAGTGAATGATGATAACATTCGTGAAACTAGAA 2642
Db 1972 GAAGAAAAATACAAGTCTCTCAACCAAGTGAATGATGATAACATTCGTGAAACTAGAA 2031
Qy 2643 CGCAGTATTTCGAGACTTGGGGAATAAATTAACCTTTTCAGTGTGTCAGACCAACACAGCA 2702
Db 2032 CGCAGTATTTCGAGACTTGGGGAATAAATTAACCTTTTCAGTGTGTCAGACCAACACAGCA 2091
Qy 2703 GGAGTACAAAATTTCAAGCTTTTGAGCAGGCTGATGAATGAATAGAGTTTCGCTTGG 2762
Db 2092 GGAGTACAAAATTTCAAGCTTTTGAGCAGGCTGATCAATGAATAGAGTTTCGCTTGG 2151
Qy 2763 ACGTACTCTCTTGATGAATCAGATGATGAATTCACATGATGAGATCCCCAGGGCAA 2822
Db 2152 ACGTACTCTCTTGATGAATCAGATGATGAATTCACATGATGAGATCCCCAGGGCAA 2211

Qy 2823 GTGTATTGCTCCCATCTTTTGACAAGAGACTCAAGCACTTCGGGTGCACAGAAGCTCTCC 2882
Db 2212 GTGTATTGCTCCCATCTTTTGACAAGAGACTCAAGCACTTCGGGTGCACAGAAGCTCTCC 2271
Qy 2883 AGTCACATTCACCTGCAAAATTTGTTGGGATACCTGTTTCCAAAGGTTTACTGGTTCAAAGA 2942
Db 2272 AGTCACATTCACCTGCAAAATTTGTTGGGATACCTGTTTCCAAAGGTTTACTGGTTCAAAGA 2331
Qy 2943 TGGGAACAGATTTCCTAAGAGAAATGAGCACTGCAAAATGAGGGGAGAAGAGATGGAC 3002
Db 2332 TGGGAACAGATTTCCTAAGAGAAATGAGCACTGCAAAATGAGGGGAGAAGAGATGGAC 2391
Qy 3003 ATGCTCTCTGCACATTCGAATCCACTACAGTGTAGAGATGGCAACTACACCATCATGGC 3062
Db 2392 ATGCTCTCTGCACATTCGAATCCACTACAGTGTAGAGATGGCAACTACACCATCATGGC 2451
Qy 3063 AGCCAACCCCCAGGGGAGAATCAGCTGTTCTGGCCACTTGTATGGTACAAAATTTGCCCAT 3122
Db 2452 AGCCAACCCCCAGGGGAGAATCAGCTGTTCTGGCCACTTGTATGGTACAAAATTTGCCCAT 2511
Qy 3123 TCGCAGTTCGGTAACTCTGCTGCTCAGTCTCAGAGGGGAAGATCCCGAGTGCNAAGAAAG 3182
Db 2512 TCGCAGTTCGGTAACTCTGCTGCTCAGTCTCAGAGGGGAAGATCCCGAGTGCNAAGAAAG 2571
Qy 3183 AGACAAAGAGCCCTCAGGAACGCTTTTCCGACCACTTTTCTCTGAGGCTCCTCTGGGA 3242
Db 2572 AGACAAAGAGCCCTCAGGAACGCTTTTCCGACCACTTTTCTCTGAGACTCCTCTGGGA 2631
Qy 3243 TATGGTAGCTCATGAGGGGCGCTCTCTCGGCTGGACTGTAAGGTGAGTGTGTACCGCC 3302
Db 2632 TATGGTAGCTCATGAGGGGCGCTCTCTCGGCTGGACTGTAAGGTGAGTGTGTACCGCC 2691
Qy 3303 CCGGAGCTGACATGGCTACTCAATGGCCAACTGCTGTCTACAGATGCTCTCCCAAGAT 3362
Db 2692 CCGGAGCTGACATGGCTACTCAATGGCCAACTGCTGTCTACAGATGCTCTCCCAAGAT 2751
Qy 3363 GCTGCTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCTCAGCCGACGC 3422
Db 2752 GCTGCTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCTCAGCCGACGC 2811
Qy 3423 AGGACCTATAAGTGTGATCGCTACCAACAAACCGGCAAGATTTCTTTAGCTGGAGCT 3482
Db 2812 AGGACCTATAAGTGTGATCGCTACCAACAAACCGGCAAGATTTCTTTAGCTGGAGCT 2871
Qy 3483 CTCTGTAGTAGCCAAAGAGGTGAAGAACCACTGTGATCCTGAGAACTACAGAACTG 3542
Db 2872 CTCTGTAGTAGCCAAAGAGGTGAAGAACCACTGTGATCCTGAGAACTACAGAACTG 2931
Qy 3543 CGGTGTTCCCGAAGGCCACCCGCTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACC 3602
Db 2932 CGGTGTTCCCGAAGGCCACCCGCTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACC 2991
Qy 3603 TGTGTTCTACTGGAAGAAAGACATGAGACATCCCTTGCACCAAGAGAGGATCAGTAT 3662
Db 2992 TGTGTTCTACTGGAAGAAAGACATGAGACATCCCTTGCACCAAGAGAGGATCAGTAT 3051
Qy 3663 GCACCAGGACACACAGGGTATGCCTGCCCTTCTCATTTACGCCAGCCAAAGAACTACAGGC 3722
Db 3052 GCACCAGGACACACAGGGTATGCCTGCCCTTCTCATTTACGCCAGCCAAAGAACTACAGGC 3111
Qy 3723 TGGATGTACACGTTGTTCAGCAAGAAATGAAGCGGATCGTGTGCTGCACTGCCAGGCT 3782
Db 3112 TGGATGTACACGTTGTTCAGCAAGAAATGAAGCGGATCGTGTGCTGCACTGCCAGGCT 3171
Qy 3783 GGATATATACGCTAAGTGGCAACCATCAGATCCCAACGCCCATGTCTGTCCGGGCCAGTGG 3842
Db 3172 GGATATATACGCTAAGTGGCAACCATCAGATCCCAACGCCCATGTCTGTCCGGGCCAGTGG 3231
Qy 3843 CAGTCCGTACGGATCTCTCACCAGTAAGGACTTGACATATTTCTGCCCTTTTCTCCTCAT 3902
Db 3232 CAGTCCGTACGGATCTCTCACCAGTAAGGACTTGACATATTTTCTGCCCTTTTCTCCTCAT 3291

QY	3903	GGAACACAGATGGTGTATTCATGCTCTTCTCGAGTGTAGTGGAGAGTGATGAACCTTTA	3962
Db	3292	GGAACACAGATGGTGTATTCATGCTCTTCTCGAGTGTAGTGGAGAGTGATGAACCTTTA	3351
QY	3963	A	3963
Db	3352	A	3352
RESULT	5		
LOCUS	AK027343		
DEFINITION	Homo sapiens cDNA FLJ14437 fis, clone HEMB1000083, weakly similar to MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES (EC 2.7.1.117).		
ACCESSION	AK027343		
VERSION	AK027343.1 GI:14041954		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens embryo, 10 weeks whole embryo, mainly body cDNA to mRNA, clone_lib:HEMB1 clone:HEMB1000083.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2127)		
AUTHORS	Isogai,T. and Otsuki,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.		
FEATURES	Location/Qualifiers		
source	1..2127		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="HEMB1000083"		
	/cissue_type="whole embryo, mainly body"		
	/clone_lib="HEMB1"		
	/dev_stage="embryo, 10 weeks"		
	/note="cloning vector: pME185FL3"		
	231..1754		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAB55048.1"		
	/db_xref="GI:14041955"		
CDS	translation="MDDSIKASTSISQLLRYSIAETRRHGNRRSRAEPPSNPCHF GSPGAAGGGGQDLDLSAFLSQELDESYNLARLAINYDPLKADETQARKRLSP DQMKLSPNLSFPNFCQDNPRSPTSKSPQAKRPOYCEQTSQKVFUNKAAEFIE LSSLFKSHSKRIRPRACKNHSKLESQNKVQENSFSDLSERRSSVPIPIPAD TRDENVNHALEQOEAKREAEQAASEAGGDTTPGSSSSLYEPLGPPPTOKLR SREVPEGRVQDCLIVGLPPQVRWYCEGKELENSPDIIHVOAGNLHSLTAEAFEE DTGRYSVFASNIYGTDSAEYIEGVSSSDSEGDPNKEMNRIQKPNRSPPTTSA VTPPAVPAQHIVAPRVATIQCCSPNYLQLDGKPIIAAPVFTKMLQNLASGEQ LVVFECRVKGADSPKRVYRECTLSDSPDFRILQKKPRMAEPCKDDFNFLLYE"		
BASE COUNT	638 a	501 c	487 g
ORIGIN	501 t		
Query Match	37.4%;	Score 1483;	DB 9; Length 2127;

		Best Local Similarity 100.0%; Pred. No. 0;		Matches 1483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGCTAT	60		
Db	231	ATCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGCTAT	290		
QY	61	TTAGCTGAAACACAGACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCCTCCAAC	120		
Db	291	TTAGCTGAAACACAGACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCCTCCAAC	350		
QY	121	CCTTGGCATTTCGGCAGTCCTTCTGGGGCGCTGGAAGAGGGCGGAGGCCAAGATGACCTT	180		
Db	351	CCTTGGCATTTCGGCAGTCCTTCTGGGGCGCTGGAAGAGGGCGGAGGCCAAGATGACCTT	410		
QY	181	CCAGATCTTTCAGCCTTCTGAGCCCAAGAAGAATTAGACCAAAAGTGTCATTTGGCAAGA	240		
Db	411	CCAGATCTTTCAGCCTTCTGAGCCCAAGAAGAATTAGACCAAAAGTGTCATTTGGCAAGA	470		
QY	241	CTGGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT	300		
Db	471	CTGGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT	530		
QY	301	TCTCTGTATCAGATGAAACACTCACCTTAATTTAAGTTTTCAGCCTAACTTCTGCCAGGAT	360		
Db	531	TCTCTGTATCAGATGAAACACTCACCTTAATTTAAGTTTTCAGCCTAACTTCTGCCAGGAT	590		
QY	361	AACCTCTGAAGTCCACCAGCTCTAAGAAGCCGCCAGGAGGCAAAAAGGCCACAGTAT	420		
Db	591	AACCTCTGAAGTCCACCAGCTCTAAGAAGCCGCCAGGAGGCAAAAAGGCCACAGTAT	650		
QY	421	TGTTCTGAAACCCAGTCCAAAAAAGTATTTTAAATAAGGCTGCCGACTTCATTGAAGAG	480		
Db	651	TGTTCTGAAACCCAGTCCAAAAAAGTATTTTAAATAAGGCTGCCGACTTCATTGAAGAG	710		
QY	481	CTATCTCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCTGCTGCAAAAAC	540		
Db	711	CTATCTCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCTGCTGCAAAAAC	770		
QY	541	CACAAGAGTAAACTGGAATCTCAAAAAGGTTATGCAAGAAAACAGCTCCAGTTTCTCA	600		
Db	771	CACAAGAGTAAACTGGAATCTCAAAAAGGTTATGCAAGAAAACAGCTCCAGTTTCTCA	830		
QY	601	GATCTGTCAAAAAGACGAGAAGATCTTCTGTTCCCATCCCTATCCCTCGGGATACCCAGG	660		
Db	831	GATCTGTCAAAAAGACGAGAAGATCTTCTGTTCCCATCCCTATCCCTCGGGATACCCAGG	890		
QY	661	GATAATGAAGTGAATCACGCCCTGGAAACACGACGAGAACCAAGAGGCGTGAAGCGGAGCAG	720		
Db	891	GATAATGAAGTGAATCACGCCCTGGAAACACGACGAGAACCAAGAGGCGTGAAGCGGAGCAG	950		
QY	721	GCTGCCAGTCAGGCGGCTGGTGAGACACTACACCAGGGTCTTCCCTTCTCTCTCTAC	780		
Db	951	GCTGCCAGTCAGGCGGCTGGTGAGACACTACACCAGGGTCTTCCCTTCTCTCTCTAC	1010		
QY	781	TATCAAGAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGGAGCAGAGAAGTT	840		
Db	1011	TATCAAGAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGGAGCAGAGAAGTT	1070		
QY	841	CCAGAAGAACTCGAGTACAGTTGGATTGATAGTGGTAGGAATTCACCACCTCAAGTA	900		
Db	1071	CCAGAAGAACTCGAGTACAGTTGGATTGATAGTGGTAGGAATTCACCACCTCAAGTA	1130		
QY	901	AGTGGTACTGTGAAGCAAGGAGCTTGAANAATCCCCCATATATTCACATCGTCCAGGCA	960		
Db	1131	AGTGGTACTGTGAAGCAAGGAGCTTGAANAATCCCCCATATATTCACATCGTCCAGGCA	1190		
QY	961	GAAATCTGCACCTCACCTGACCATTTGCGGAAGCCCTTTGAAGAGGACACAGACGCTATTC	1020		
Db	1191	GAAATCTGCACCTCACCTGACCATTTGCGGAAGCCCTTTGAAGAGGACACAGACGCTATTC	1250		
QY	1021	TGCTTTGCTTCTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG	1080		

Db 1251 TGCTTTGCTTCTAATCATATGAGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1310
Qy 1081 GTTCTCTTCTGACTCAGAAAGGACCTTAACAAGGAGAGATGAATCGAATCCAGAG 1140
Db 1311 GTTCTCTTCTGACTCAGAAAGGACCTTAACAAGGAGAGATGAATCGAATCCAGAG 1370
Qy 1141 CCAATAGAGGTGCATCTCTCCCTCCACTACCTCTCTGAGTTCCTCCAGCAGTACCCCAA 1200
Db 1371 CCAATAGAGGTGCATCTCTCCCTCCACTACCTCTCTGAGTTCCTCCAGCAGTACCCCAA 1430
Qy 1201 GCCAGCATTTGGTGGCCCAACCTCGTGTGCAACCATCCAGAGTGTGAGAGCCCCACC 1260
Db 1431 GCCAGCATTTGGTGGCCCAACCTCGTGTGCAACCATCCAGAGTGTGAGAGCCCCACC 1490
Qy 1261 AATTACTTGCAGGATGAGTGGAAACCTATCATTTGCAGCTCTCTGTTTGTACAAAGATG 1320
Db 1491 AATTACTTGCAGGATGAGTGGAAACCTATCATTTGCAGCTCTCTGTTTGTACAAAGATG 1550
Qy 1321 CTACAAAATTTGTGAGCTTCTGAGGGTCAGCTGGTGTGTTTGAATGCAGAGTAAAGGA 1380
Db 1551 CTACAAAATTTGTGAGCTTCTGAGGGTCAGCTGGTGTGTTTGAATGCAGAGTAAAGGA 1610
Qy 1381 GCTCCATCTCTAAGTGTGAGTGTATAGAGAGGGACTTTAATAGAGATTTCTCCAGAT 1440
Db 1611 GCTCCATCTCTAAGTGTGAGTGTATAGAGAGGGACTTTAATAGAGATTTCTCCAGAT 1670
Qy 1441 TTTAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCAG 1483
Db 1671 TTTAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCAG 1713

RESULT 6
AC024258 160658 bp DNA linear PRI 14-MAY-2002
LOCUS Homo sapiens chromosome 10 clone RP11-297N15, complete sequence.
AC024258
VERSION AC024258.9 GI:20564414
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith,D.R.
REFERENCE Smith,D.R.
AUTHORS Genome Therapeutics Corporation Sequencing Center: Human Genome
TITLE Sequence Data
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 160658)
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 160658)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 160658)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 14, 2002 this sequence version replaced gi:16905139.
FEATURES
source Location/Qualifiers
1..160658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-297N15"
/clone_lib="RC1-11"

BASE COUNT 50744 a 31675 c 30448 g 47791 t
ORIGIN

Query Match 22.8%; Score 904; DB 9; Length 160658;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAAGACGACAGATAGAAAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 60
Db 110106 ATGCAAGACGACAGATAGAAAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 110165
Qy 61 TTAGCTGAAACCCAGACATCGGGGAAACAATGAGAGGAGTTCGAGCGGAGCCCTCTCCCAAC 120
Db 110166 TTAGCTGAAACCCAGACATCGGGGAAACAATGAGAGGAGTTCGAGCGGAGCCCTCTCCCAAC 110225
Qy 121 CTTTGCATTTTCGGCAGTCTCTTCTGGGGCCGCTGAAGGAGGCGGAGGCAAGATGACCTT 180
Db 110226 CTTTGCATTTTCGGCAGTCTCTTCTGGGGCCGCTGAAGGAGGCGGAGGCAAGATGACCTT 110285
Qy 181 CCAGATCTTTCAGCCTTTCTGAGCCAAAGAAATTTAGACGAAAGTGCAATTTGGCAAGA 240
Db 110286 CCAGATCTTTCAGCCTTTCTGAGCCAAAGAAATTTAGACGAAAGTGCAATTTGGCAAGA 110345
Qy 241 CTGGCCATCAATTTACGACCTTTTGGAGAAGGCAGATGAACTCAAGCTAGAAAAACGACTT 300
Db 110346 CTGGCCATCAATTTACGACCTTTTGGAGAAGGCAGATGAACTCAAGCTAGAAAAACGACTT 110405
Qy 301 TCTCTGATCAGATGAACACTCACCTTAATTTAAGTTTTTGAGCCTTAACCTTCTGCCAGAT 360
Db 110406 TCTCTGATCAGATGAACACTCACCTTAATTTAAGTTTTTGAGCCTTAACCTTCTGCCAGAT 110465
Qy 361 AACCTCGAAGTCCACAGCTCTAAAGAAAGCCCCCAGAGGCAAAAAGGCCACAGTAT 420
Db 110466 AACCTCGAAGTCCACAGCTCTAAAGAAAGCCCCCAGAGGCAAAAAGGCCACAGTAT 110525
Qy 421 TGTTCGAAACCCAGTCCAAAAAGATATTTTAAATAGGCTCGCGACTTCATTTGAAGAG 480
Db 110526 TGTTCGAAACCCAGTCCAAAAAGATATTTTAAATAGGCTCGCGACTTCATTTGAAGAG 110585
Qy 481 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGCTGCGTCAAAAAC 540
Db 110586 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGCTGCGTCAAAAAC 110645
Qy 541 CACAAGAGTAACCTGGGAATCTCAAAACAAGTTTATGAGGAAAAACAGCTCCAGTTTCTCA 600
Db 110646 CACAAGAGTAACCTGGGAATCTCAAAACAAGTTTATGAGGAAAAACAGCTCCAGTTTCTCA 110705
Qy 601 GATCTGTCAAGAACGACGAGAAAGATCTTCTGTCCCATCCCTATCCCTGCGGATACACAG 660
Db 110706 GATCTGTCAAGAACGACGAGAAAGATCTTCTGTCCCATCCCTATCCCTGCGGATACACAG 110765
Qy 661 GATAATGAAGTGAATCAGCCCTCGAAACAGCAGAGGAAAGCCGTTGAAGCGGAGCAG 720
Db 110766 GATAATGAAGTGAATCAGCCCTCGAAACAGCAGAGGAAAGCCGTTGAAGCGGAGCAG 110825
Qy 721 GCTGCCAGTGAAGCGGCTGTGGAGACACTACACAGGGTCTTCCCTTCATCTCTGTAC 780
Db 110826 GCTGCCAGTGAAGCGGCTGTGGAGACACTACACAGGGTCTTCCCTTCATCTCTGTAC 110885
Qy 781 TATGAAGAACCTCTGGGGCAACCTCCCGGTTCACTCAAAAAGTTACGAGCAGAGAGAGTT 840
Db 110886 TATGAAGAACCTCTGGGGCAACCTCCCGGTTCACTCAAAAAGTTACGAGCAGAGAGAGTT 110945
Qy 841 CCAGAAGGAACCTCGAGTACAGTTGGATTGTCATAGTGTAGGAATTCACCAACCTCAAGTA 900
Db 110946 CCAGAAGGAACCTCGAGTACAGTTGGATTGTCATAGTGTAGGAATTCACCAACCTCAAGTA 111005
Qy 901 AGGT 904
Db 111006 AGGT 111009

RESULT 7
AC063964 175265 bp DNA linear HTG 30-AUG-2001
LOCUS Homo sapiens chromosome 10 clone RP11-31303, WORKING DRAFT
DEFINITION

```
SEQUENCE, 20 unordered pieces.
AC063964
VERSION AC063964.3 GI:9887616
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 175265)
JOURNAL Smith,D.R.
AUTHORS Genome Therapeutics Corporation Sequencing Center: Human Genome
TITLE Sequence Data
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 175265)
TITLE Smith,D.R.
JOURNAL Direct Submission
COMMENT Submitted (22-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Aug 24, 2000 this sequence version replaced gi:8439815.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg198
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 153700 bases at least Q40
Consensus quality: 164398 bases at least Q30
Consensus quality: 167130 bases at least Q20
Insert size: 173464; sum-of-contigs
Quality coverage: 4.1x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1145: contig of 1145 bp in length
* 1146 1245: gap of unknown length
* 1246 2414: contig of 1169 bp in length
* 2415 2514: gap of unknown length
* 2515 3864: contig of 1350 bp in length
* 3865 3964: gap of unknown length
* 3965 5078: contig of 1114 bp in length
* 5079 5178: gap of unknown length
* 5179 6647: contig of 1469 bp in length
* 6648 6747: gap of unknown length
* 6748 7926: contig of 1179 bp in length
* 7927 8027: gap of unknown length
* 8028 9412: contig of 1386 bp in length
* 9413 9512: gap of unknown length
* 9513 10563: contig of 1051 bp in length
* 10564 10663: gap of unknown length
* 10664 11995: contig of 1332 bp in length
* 11996 12095: gap of unknown length
* 12096 13309: contig of 1214 bp in length
* 13310 13409: gap of unknown length
* 13410 15172: contig of 1763 bp in length
* 15173 15272: gap of unknown length
* 15273 18073: contig of 2701 bp in length
* 17974 18074: gap of unknown length
* 18074 21419: contig of 3345 bp in length
* 21420 24881: gap of unknown length
* 24882 24981: gap of unknown length
* 24982 31876: contig of 6895 bp in length
* 31877 31976: gap of unknown length
* 31977 49015: contig of 17039 bp in length
* 49016 49115: gap of unknown length
* 49116 66287: contig of 17171 bp in length
* 66288 66387: gap of unknown length
* 66388 86201: contig of 19815 bp in length
* 86202 86302: gap of unknown length
* 86303 132480: contig of 46179 bp in length
* 132481 132581: gap of unknown length
* 132582 175265: contig of 42685 bp in length.
FEATURES
Source
Location/Qualifiers
1..175265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-31303"
/clone_lib="RPC1-11"
1..1145
/note="assembly_name:Contig1"
misc_feature 1246..2414
/note="assembly_name:Contig11"
misc_feature 2515..3864
/note="assembly_name:Contig12"
misc_feature 3965..5078
/note="assembly_name:Contig15"
misc_feature 5179..6647
/note="assembly_name:Contig18"
misc_feature 6748..7926
/note="assembly_name:Contig20"
misc_feature 8027..9412
/note="assembly_name:Contig23"
misc_feature 9513..10563
/note="assembly_name:Contig25"
misc_feature 10664..11995
/note="assembly_name:Contig27"
misc_feature 12096..13309
/note="assembly_name:Contig28"
misc_feature 13410..15172
/note="assembly_name:Contig29"
misc_feature 15273..17973
/note="assembly_name:Contig30"
misc_feature 18074..21419
/note="assembly_name:Contig31"
misc_feature 21520..24881
/note="assembly_name:Contig32"
misc_feature 24982..31876
/note="assembly_name:Contig33"
misc_feature 31977..49015
/note="assembly_name:Contig34"
misc_feature 49116..66286
/note="assembly_name:Contig35"
clone_end:SP6"
misc_feature 66387..86201
/note="assembly_name:Contig36"
misc_feature 86302..132480
/note="assembly_name:Contig37"
clone_end:T7"
misc_feature 132581..175265
/note="assembly_name:Contig38"
BASE COUNT 53132 a 34451 c 33650 g 52102 t 1930 others
ORIGIN
Query Match 22.8%; Score 904; DB 2; Length 175265;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAAGACGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 60
|||||
Db 162072 ATGCAAGACGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 162131
QY 61 TTAGCTGAAACCCAGACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCTCCCAAC 120
|||||
Db 162132 TTAGCTGAAACCCAGACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCTCCCAAC 162191
```

```
Oy 121 CTTGCCATTTTCGCGAGTCTTCTGGGCCGCTGAAGGAGCGGAGGCGCAAGATGACCTT 180
|||||
Db 162192 CTTGCCATTTTCGCGAGTCTTCTGGGCCGCTGAAGGAGCGGAGGCGCAAGATGACCTT 162251
|||||
Oy 181 CCAGATCTTTTCAGGCTTTCTGAGCAGAGAGAAATAGAGAAAGTGCAATTTGGCAGA 240
|||||
Db 162252 CCAGATCTTTTCAGGCTTTCTGAGCAGAGAGAAATAGAGAAAGTGCAATTTGGCAGA 162311
|||||
Oy 241 CTGGCCATCAATTAGGACCTTTTGGAGAGGCGAGATGAATCAAGCTAGAAAAAGGACTT 300
|||||
Db 162312 CTGGCCATCAATTAGGACCTTTTGGAGAGGCGAGATGAATCAAGCTAGAAAAAGGACTT 162371
|||||
Oy 301 TCTCCTGATCAGATGAACACTCACCTCAATTAAGTTTGTAGCCCTAAGTCTGCGCAGGAT 360
|||||
Db 162372 TCTCCTGATCAGATGAACACTCACCTCAATTAAGTTTGTAGCCCTAAGTCTGCGCAGGAT 162431
|||||
Oy 361 AACCTCTGAAGTCCGACAGCTCTAAGAGAGGCCCGCCAGGAGGCAAAAGGCCACAGTAT 420
|||||
Db 162432 AACCTCTGAAGTCCGACAGCTCTAAGAGAGGCCCGCCAGGAGGCAAAAGGCCACAGTAT 162491
|||||
Oy 421 TGTTCCTGAACCCAGTCCAAAAAGTATTTTAAATAGGCTGCCGACTTTCATTGAAGAG 480
|||||
Db 162492 TGTTCCTGAACCCAGTCCAAAAAGTATTTTAAATAGGCTGCCGACTTTCATTGAAGAG 162551
|||||
Oy 481 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAAGGATTAGACCTCGTGCCTGCAAAAAC 540
|||||
Db 162552 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAAGGATTAGACCTCGTGCCTGCAAAAAC 162611
|||||
Oy 541 CACAAGAGTAACTGGAACTCTCAAAACAAGATTATGAGGAGGAAACAGTCCAGTTCTCTCA 600
|||||
Db 162612 CACAAGAGTAACTGGAACTCTCAAAACAAGATTATGAGGAGGAAACAGTCCAGTTCTCTCA 162671
|||||
Oy 601 GATCTGTGAGAAACAGAGAAAGATCTTCTGTCCCATCCCTATCCCTGCGGATACACAGG 660
|||||
Db 162672 GATCTGTGAGAAACAGAGAAAGATCTTCTGTCCCATCCCTATCCCTGCGGATACACAGG 162731
|||||
Oy 661 GATAATGAAGTGAATCACGCGCTCGAAGCAGCAGGAGGAGGCGGTGAAGCGGAGCAG 720
|||||
Db 162732 GATAATGAAGTGAATCACGCGCTCGAAGCAGCAGGAGGAGGCGGTGAAGCGGAGCAG 162791
|||||
Oy 721 GCTGCCAGTAGGCGGCTGTGGAGACACTACACAGGAGTCTTCCCTTCATCTCTCTAC 780
|||||
Db 162792 GCTGCCAGTAGGCGGCTGTGGAGACACTACACAGGAGTCTTCCCTTCATCTCTCTAC 162851
|||||
Oy 781 TATGAAGAACTCTGGGGCAACCTCCCGGTTCACTCAAAAGTTTACGAGCAGAGAAAGTT 840
|||||
Db 162852 TATGAAGAACTCTGGGGCAACCTCCCGGTTCACTCAAAAGTTTACGAGCAGAGAAAGTT 162911
|||||
Oy 841 CCAGAAGAACTCGAGTACAGTTGGATTGTCATAGTGTAGGAATTTCCACCACCTCAAGTA 900
|||||
Db 162912 CCAGAAGAACTCGAGTACAGTTGGATTGTCATAGTGTAGGAATTTCCACCACCTCAAGTA 162971
|||||
Oy 901 AGGT 904
|||||
Db 162972 AGGT 162975
|||||

RESULT 8
AC024460/c
LOCUS AC024460 180707 bp DNA linear HTG 28-MAR-2000
DEFINITION Homo sapiens clone RP11-287P14, WORKING DRAFT SEQUENCE, 22
unordered pieces.
ACCESSION AC024460
VERSION AC024460.2 GI:7331497
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180707)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-287P14
```

JOURNAL
REFERENCE
AUTHORSUnpublished
2 (bases 1 to 180707)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavskiy,L.,
Boukigater,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grant,Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,
Pieterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 28, 2000 this sequence version replaced gi:7108256.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7059

Center clone name: 287_P_14

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 168456 bases at least Q40

Consensus quality: 174769 bases at least Q30

Consensus quality: 176914 bases at least Q20

Insert size: 188000; agarose-fp

Insert size: 178607; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 1528: contig of 1528 bp in length
* 1529 1628: gap of 100 bp
* 1629 2914: contig of 1286 bp in length
* 2915 3014: gap of 100 bp
* 3015 5061: contig of 2047 bp in length
* 5062 5161: gap of 100 bp
* 5162 7974: contig of 2813 bp in length
* 7975 8074: gap of 100 bp
* 8075 11732: contig of 3658 bp in length
* 11733 11832: gap of 100 bp
* 11833 14787: contig of 2955 bp in length
* 14788 14887: gap of 100 bp
* 14888 19746: contig of 4859 bp in length
* 19747 19846: gap of 100 bp
* 19847 23870: contig of 4024 bp in length

```
* 23871 23970: gap of 100 bp
* 23971 28748: contig of 4778 bp in length
* 28749 28848: gap of 100 bp
* 28849 35292: contig of 6444 bp in length
* 35293 35392: gap of 100 bp
* 35393 40557: contig of 4965 bp in length
* 40558 40457: gap of 100 bp
* 40458 45936: contig of 5479 bp in length
* 45937 46036: gap of 100 bp
* 46037 52180: contig of 6144 bp in length
* 52181 52280: gap of 100 bp
* 52281 57511: contig of 5231 bp in length
* 57512 57611: gap of 100 bp
* 57612 65019: contig of 7408 bp in length
* 65020 65119: gap of 100 bp
* 65120 73479: contig of 8360 bp in length
* 73480 73579: gap of 100 bp
* 73580 81893: contig of 8314 bp in length
* 81894 81993: gap of 100 bp
* 81994 92063: contig of 10070 bp in length
* 92064 92163: gap of 100 bp
* 92164 103396: contig of 11233 bp in length
* 103397 103496: gap of 100 bp
* 103497 112935: contig of 9439 bp in length
* 112936 113035: gap of 100 bp
* 113036 126895: contig of 13860 bp in length
* 126896 126995: gap of 100 bp
* 126996 180707: contig of 53712 bp in length.
FEATURES
    source
    1. .180707
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="Rp11-287P14"
        /clone_lib="RPC1-11 Human Male BAC"
    1. .1528
        /note="assembly_fragment"
        1629. .2914
        /note="assembly_fragment"
        3015. .5061
        /note="assembly_fragment"
        5162. .7974
        /note="assembly_fragment"
        8075. .11732
        /note="assembly_fragment"
        11833. .14787
        /note="assembly_fragment"
        14888. .19746
        /note="assembly_fragment"
        19847. .23870
        /note="assembly_fragment"
        23971. .28748
        /note="assembly_fragment"
        28849. .35292
        /note="assembly_fragment"
        35393. .40357
        /note="assembly_fragment"
        40458. .45936
        /note="assembly_fragment"
        46037. .52180
        /note="assembly_fragment"
        52281. .57511
        /note="assembly_fragment"
        57612. .65019
        /note="assembly_fragment"
        65120. .73479
        /note="assembly_fragment"
        73580. .81893
        /note="assembly_fragment"
        81994. .92063
        /note="assembly_fragment"
        92164. .103396
        /note="assembly_fragment"
        103497. .112935
```

```
/note="assembly_fragment"
clone_end:T7
vector_side:left"
113036..126895
/note="assembly_fragment"
126996..180707
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
BASE COUNT 55951 a 33545 c 34680 g 54429 t 2102 others
ORIGIN
Query Match 22.8%; Score 904; DB 2; Length 180707;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 60
|||||
Db 154380 ATGCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 154321
|||||
QY 61 TTAGCTGAACACGACATCGGGGAAACAATGAGAGAGTGGAGGAGCCCTCTCTCCAAC 120
|||||
Db 154320 TTAGCTGAACACGACATCGGGGAAACAATGAGAGAGTGGAGGAGCCCTCTCTCCAAC 154261
|||||
QY 121 CTTTGGCATTTCGGCAGTCTTCTTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 180
|||||
Db 154260 CTTTGGCATTTCGGCAGTCTTCTTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 154201
|||||
QY 181 CCAGATCTTTGAGCCTTTCTGAGCCAAAGAAGATTAGACGAAAAGTGTCATTTGGCAAGA 240
|||||
Db 154200 CCAGATCTTTGAGCCTTTCTGAGCCAAAGAAGATTAGACGAAAAGTGTCATTTGGCAAGA 154141
|||||
QY 241 CTGGCCATCAATTACGACCCCTTTGGAGAGGCGAGATCAAACTCAAGCTAGAAAACGACTT 300
|||||
Db 154140 CTGGCCATCAATTACGACCCCTTTGGAGAGGCGAGATCAAACTCAAGCTAGAAAACGACTT 154081
|||||
QY 301 TCTCCTGATCAGATGAACACACTCACCTTAATTTAAGTTTTGAGCCTAACTTCTTGCAGGAT 360
|||||
Db 154080 TCTCCTGATCAGATGAACACACTCACCTTAATTTAAGTTTTGAGCCTAACTTCTTGCAGGAT 154021
|||||
QY 361 AACCTCGAAGTCCACCAGCTCTAAAGAAGCCGCCAGGAGGCAAAAGGCCACAGTAT 420
|||||
Db 154020 AACCTCGAAGTCCACCAGCTCTAAAGAAGCCGCCAGGAGGCAAAAGGCCACAGTAT 153961
|||||
QY 421 TGTCTGAACCCAGTCCAAAAGATTTTAAATAAGCTGCCGACTTCATTGAAGAG 480
|||||
Db 153960 TGTCTGAACCCAGTCCAAAAGATTTTAAATAAGCTGCCGACTTCATTGAAGAG 153901
|||||
QY 481 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGCTGCTGCAAAAAC 540
|||||
Db 153900 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGCTGCTGCAAAAAC 153841
|||||
QY 541 CACAAGAGTAACTGGAAATCTCAAAACAAGTTATGACGAGAAAACAGCTCCAGTTTCTCA 600
|||||
Db 153840 CACAAGAGTAACTGGAAATCTCAAAACAAGTTATGACGAGAAAACAGCTCCAGTTTCTCA 153781
|||||
QY 601 GATCTGTCAAAAGACGAGAAAGATCTTCTTCCCATCCCTATCCCTCGGATACACAGG 660
|||||
Db 153780 GATCTGTCAAAAGACGAGAAAGATCTTCTTCCCATCCCTATCCCTCGGATACACAGG 153721
|||||
QY 661 GATAATGAAGTGAATCAGCCCTCGGAACAGCAGGAAGCCGAAGGCGTGAAGCGGAGCAG 720
|||||
Db 153720 GATAATGAAGTGAATCAGCCCTCGGAACAGCAGGAAGCCGAAGGCGTGAAGCGGAGCAG 153661
|||||
QY 721 GCTGCCAGTGAGGCGCTGGTGGAGACACTACACAGGGTCTTCCCTTCACTCTGTAC 780
|||||
Db 153660 GCTGCCAGTGAGGCGCTGGTGGAGACACTACACAGGGTCTTCCCTTCACTCTGTAC 153601
|||||
QY 781 TATGAACAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTACGAGCAGAGAGATT 840
|||||
Db 153600 TATGAACAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTACGAGCAGAGAGATT 153541
|||||
QY 841 CCAGAAGGAACTCGAGTACAGTTGGATTGTCATAGTGGTAGGAATTTCCACCACCTCAAGTA 900
```

```
|||||
Db 153540 CCAGAAGAACTGAGTACAGTTGGTAGTGGTAGGAATTCACCACCTCAAGTA 153481
Qy 901 AAGT 904
|||||
Db 153480 AAGT 153477

RESULT 9
AL512429 11170 bp DNA linear PRI 26-OCT-2001
LOCUS Human DNA sequence from clone RP11-429C3 on chromosome 10, complete
DEFINITION sequence.
ACCESSION AL512429
VERSION AL512429.14 GI:16508239
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 11170)
AUTHORS Mashreghi-Mohammadi, M.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Oct 29, 2001 this sequence version replaced gi:15028699.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-429C3 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-429C3. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-153K11 is at 9171 in this sequence.
The true right end of clone RP11-297N15 is at 2000 in this
sequence.

FEATURES
source Location/Qualifiers
1. .11170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-429C3"
/clone_lib="RPCI-11.2"
1620. .2069
/note="match: GSS: Em:AQ435766"
complement(2076..2585)
/note="match: GSS: Em:AQ822926"
2596. .3437
/note="match: GSS: Em:AQ746971"
3298. .3591
misc_feature
/note="match: GSS: Em:AQ435766"
misc_feature
/note="match: GSS: Em:AQ822926"
misc_feature
/note="match: GSS: Em:AQ746971"
misc_feature
/note="match: GSS: Em:AQ435766"
```

```
/note="match: GSS: Em:AQ742982"
3596. .3953
/note="match: GSS: Em:AQ742982"
complement(7993..8507)
/note="match: GSS: Em:AQ881403"
9218. .9756
/note="match: GSS: Em:AQ386388"
BASE COUNT 3230 a 2428 c 2215 g 3297 t
ORIGIN

Query Match 10.4%; Score 413; DB 9; Length 11170;
Best Local Similarity 99.5%; Pred. No. 5.9e-218;
Matches 563; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1974 TGATTCCACTCAGTTACAACAGCTTCATACCAAGTCTTACTGGAACAACACCAATTCGA 2033
Db 4075 TGATTCCACTCAGTTACAACAGCTTCATACCAAGTCTTACTGGAACAACACCAATTCGA 4134
Qy 2034 AAACCCACCTCCTTCATCTCCTTAAGAGGATTTCCCTTCARCATGACTGTTTGAACCTCAA 2093
Db 4135 AAACCCACCTCCTTCATCTCCTTAAGAGGATTTCCCTTCAGCATGACTGTTTGAACCTCAA 4194
Qy 2094 TGCTCCCGCCGCGGTGACACATCCATTAAGCAGGTGAAGGCTCCTTCATCACAGAGGTT 2153
Db 4195 TGCTCCCGCCGCGGTGACACATCCATTAAGCAGGTGAAGGCTCCTTCATCACAGAGGTT 4254
Qy 2154 CAGCTTGGCCGCGCGGAAGTATTTCTCCCTCCACCAACACACCGCAGCACTGTGGC 2213
Db 4255 CAGCTTGGCCGCGCGGAAGTATTTCTCCCTCCACCAACACACCGCAGCACTGTGGC 4314
Qy 2214 CCCTTCCAGCTCTCCGGTGTTCACCTTTGAGCAGCACCTCCTCAAACTATTTCAGAGGACAGT 2273
Db 4315 CCCTTCCAGCTCTCCGGTGTTCACCTTTGAGCAGCACCTCCTCAAACTATTTCAGAGGACAGT 4374
Qy 2274 GAGCAAGAAGGCTCTTAGTGTCTCACCCCTCTGTGCAAACTAAATTCCTCAGAGGGCT 2333
Db 4375 GAGCAAGAAGGCTCTTAGTGTCTCACCCCTCTGTGCAAACTAAATTCCTCAGAGGGCT 4434
Qy 2334 TTCCATCAAAATGAGCCACTCCACCAAGGCGCCCAAGACACCAACACCACTTCAC 2393
Db 4435 TTCCATCAAAATGAGCCACTCCACCAAGGCGCCCAAGACACCAACACCACTTCAC 4494
Qy 2394 ATTTTCCATCCCGAGCGGAACACAGTTTCAGCCCGCGTGTGTGTCGCCAATTCCTGCTC 2453
Db 4495 ATTTTCCATCCCGAGCGGAACACAGTTTCAGCCCGCGTGTGTGTCGCCAATTCCTGCTC 4554
Qy 2454 TCCTACAGCGCGGATTCAGAACCCAGTGGCTTTCCTCAGCTCTGTTCTGCTCTCTCC 2513
Db 4555 TCCTACAGCGCGGATTCAGAACCCAGTGGCTTTCCTCAGCTCTGTTCTGCTCTCTCC 4614
Qy 2514 TGCCATCCCGACCCACAAATGCCATGG 2539
Db 4615 TGCCATCCCGACCCACAAATGCCATGG 4640

RESULT 10
AL442068 290625 bp DNA linear HTG 20-NOV-2001
LOCUS Homo sapiens chromosome 10 clone RP11-331G8, *** SEQUENCING IN
DEFINITION PROGRESS ***, 43 unordered pieces.
ACCESSION AL442068
VERSION AL442068.16 GI:17043795
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Sims, S.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

COMMENT

On Nov 21, 2001 this sequence version replaced gi:14575284.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA331G8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: pIasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 277809 bases at least Q40
Consensus quality: 281712 bases at least Q30
Consensus quality: 284299 bases at least Q20
Insert size: 286425; sum-of-contigs
Insert size: 156064; 7.4% error; agarose-fp
Quality coverage: 5.79x in Q20 bases; sum-of-contigs Quality
coverage: 15.47x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 103320: contig of 103320 bp in length
* 103321 103420: gap of 100 bp
* 103421 140724: contig of 37304 bp in length
* 140725 140824: gap of 100 bp
* 140825 162687: contig of 21863 bp in length
* 162688 162787: gap of 100 bp
* 162788 165269: contig of 2482 bp in length
* 165270 165369: gap of 100 bp
* 165370 167787: contig of 2418 bp in length
* 167788 167887: gap of 100 bp
* 167888 170672: contig of 2785 bp in length
* 170673 170772: gap of 100 bp
* 170773 173340: contig of 2568 bp in length
* 173341 173440: gap of 100 bp
* 173441 176369: contig of 2929 bp in length
* 176370 176469: gap of 100 bp
* 176470 179660: contig of 3191 bp in length
* 179661 179760: gap of 100 bp
* 179761 181788: contig of 2028 bp in length
* 181789 181888: gap of 100 bp
* 181889 185591: contig of 3703 bp in length
* 185592 185691: gap of 100 bp
* 185692 187897: contig of 2206 bp in length
* 187898 187997: gap of 100 bp
* 187998 190151: contig of 2134 bp in length
* 190152 190251: gap of 100 bp
* 190252 192977: contig of 2726 bp in length
* 192978 193077: gap of 100 bp
* 193078 195504: contig of 2427 bp in length
* 195505 195604: gap of 100 bp
* 195605 198738: contig of 3134 bp in length
* 198739 198838: gap of 100 bp
* 198839 202505: contig of 3667 bp in length
* 202506 202605: gap of 100 bp
* 202606 206757: contig of 4152 bp in length
* 206758 208857: gap of 100 bp
* 208858 211910: contig of 5053 bp in length
* 211911 212010: gap of 100 bp
* 212011 218152: contig of 6142 bp in length
* 218153 218252: gap of 100 bp
* 218253 221882: contig of 3630 bp in length
* 221883 221982: gap of 100 bp
* 221983 225735: contig of 3753 bp in length
* 225736 225835: gap of 100 bp
* 225836 229907: contig of 4072 bp in length

* 229908 230007: gap of 100 bp
* 230008 233016: contig of 3009 bp in length
* 233017 233116: gap of 100 bp
* 233117 236145: contig of 3029 bp in length
* 236146 236245: gap of 100 bp
* 236246 239609: contig of 3364 bp in length
* 239610 239709: gap of 100 bp
* 239710 241796: contig of 2087 bp in length
* 241797 241896: gap of 100 bp
* 241897 243915: contig of 2019 bp in length
* 243916 244015: gap of 100 bp
* 244016 246442: contig of 2427 bp in length
* 246443 246542: gap of 100 bp
* 246543 249599: contig of 3057 bp in length
* 249600 249699: gap of 100 bp
* 249700 252680: contig of 2981 bp in length
* 252681 252780: gap of 100 bp
* 252781 256312: contig of 3532 bp in length
* 256313 256412: gap of 100 bp
* 256413 259563: contig of 3151 bp in length
* 259564 259663: gap of 100 bp
* 259664 262204: contig of 2541 bp in length
* 262205 262304: gap of 100 bp
* 262305 264799: contig of 2495 bp in length
* 264800 264899: gap of 100 bp
* 264900 267645: contig of 2746 bp in length
* 267646 267745: gap of 100 bp
* 267746 271525: contig of 3780 bp in length
* 271526 271625: gap of 100 bp
* 271626 275798: contig of 4173 bp in length
* 275799 275898: gap of 100 bp
* 275899 278025: contig of 2127 bp in length
* 278026 278125: gap of 100 bp
* 278126 281433: contig of 3308 bp in length
* 281434 281533: gap of 100 bp
* 281534 284437: contig of 2904 bp in length
* 284438 284537: gap of 100 bp
* 284538 286727: contig of 2190 bp in length
* 286728 286827: gap of 100 bp
* 286828 290625: contig of 3798 bp in length.

FEATURES
 Source
 1..290625
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone_lib="RPCI-11.2"
 1..103320
 /notes="assembly_fragment:07950
 fragment_chain:1
 clone_end:T7
 vector_side:left"
 103421..140724
 /notes="assembly_fragment:01658
 fragment_chain:1"
 140825..162687
 /notes="assembly_fragment:01866
 fragment_chain:1"
 162788..165269
 /notes="assembly_fragment:01933
 fragment_chain:2"
 165370..167787
 /notes="assembly_fragment:02386
 fragment_chain:2"
 167888..170672
 /notes="assembly_fragment:02267
 fragment_chain:3"
 170773..173340
 /notes="assembly_fragment:02094
 fragment_chain:3"
 173441..176369
 /notes="assembly_fragment:00062"
 176470..179660

misc_feature	/note="assembly_fragment:00157" 179761..181788				
misc_feature	/note="assembly_fragment:00471" 181889..185591				
misc_feature	/note="assembly_fragment:00485" 185692..187897				
misc_feature	/note="assembly_fragment:00607" 187998..190151				
misc_feature	/note="assembly_fragment:01019" 190252..192977				
misc_feature	/note="assembly_fragment:01432" 193078..195504				
misc_feature	/note="assembly_fragment:01879" 195605..198738				
misc_feature	/note="assembly_fragment:01944" 198839..202505				
misc_feature	/note="assembly_fragment:02174" 202606..206757				
misc_feature	/note="assembly_fragment:02200" 206858..211910				
misc_feature	/note="assembly_fragment:02684" 212011..218152				
misc_feature	/note="assembly_fragment:03379" 218253..221882				
misc_feature	/note="assembly_fragment:03527" 221983..225735				
misc_feature	/note="assembly_fragment:03880" 225836..229907				
misc_feature	/note="assembly_fragment:04098" 230008..233016				
misc_feature	/note="assembly_fragment:04166" 233117..236145				
misc_feature	/note="assembly_fragment:04570" 236246..239609				
misc_feature	/note="assembly_fragment:04709" 239710..241796				
misc_feature	/note="assembly_fragment:04939" 241897..243915				
Query Match 10.4%; Score 413; DB 2; Length 290625;					
Best Local Similarity 99.5%; Pred. No. 4.3e-218;					
Matches 563; Conservative 0; Mismatches 3; Indels 0; Gaps 0;					
QY 1974	TGATTCACCTCAGTTACACAGCTTCATAACCAAGTCTTACTGGAAACACCAATTCGA	2033			
Db 213297	TGATTCACCTCAGTTACACAGCTTCATAACCAAGTCTTACTGGAAACACCAATTCGA	213356			
QY 2034	AAACCCACCTCCTTCATCTCCTTAAGGAGTTCTCTTCARCATGACTGTTTGAACCTCAA	2093			
Db 213357	AAACCCACCTCCTTCATCTCCTTAAGGAGTTCTCTTCAGCATGACTGTTTGAACCTCAA	213416			
QY 2094	TGCTCCCCAGCGGTGACAACTCATTAAGCAGGTGAAGGCTCCTTCATCACAGAGTT	2153			
Db 213417	TGCTCCCCAGCGGTGACAACTCATTAAGCAGGTGAAGGCTCCTTCATCACAGAGTT	213476			
QY 2154	CAGCTTGCCCGCGCGGAAGTATTTCTCCCTCCACGAACACCCAGCAACTGTGGC	2213			
Db 213477	CAGCTTGCCCGCGCGGAAGTATTTCTCCCTCCACGAACACCCAGCAACTGTGGC	213536			
QY 2214	CCCTTCAGCTCTCCGGTGTTCACCTTTGAGCAGCACTCTCAAACTATTGAGGACAGT	2273			
Db 213537	CCCTTCAGCTCTCCGGTGTTCACCTTTGAGCAGCACTCTCAAACTATTGAGGACAGT	213596			
QY 2274	GAGCAAGAAAGCCTCTTAGTGCTCTACCCCTCTGTGCAACCAAAATCTCCAGGAGGCT	2333			
Db 213597	GAGCAAGAAAGCCTCTTAGTGCTCTACCCCTCTGTGCAACCAAAATCTCCAGGAGGCT	213656			
QY 2334	TTCCATCCAAATGAGCCACTCCACACAGGCCCCCAACAGAACCCACCAACCAATTCAC	2393			
Db 213657	TTCCATCCAAATGAGCCACTCCACACAGGCCCCCAACAGAACCCACCAACCAATTCAC	213716			
QY 2394	ATTTTCCATCCCGAGGGAACCAAGTTTCAGCCCGCTGTGTGTCGCCCAATTCCTGTCTC	2453			

Db 213717	ATTTTCCATCCCGAGGGAACCAAGTTTCAGCCCGCTGTGTGTCGCCCAATTCCTGTCTC	213776			
QY 2454	TCCTACCAGCGGATTCAGAACCCAGTGGCTTCTCAGCTCTGCTTCTGCTTCTCTCCC	2513			
Db 213777	TCCTACCAGCGGATTCAGAACCCAGTGGCTTCTCAGCTCTGCTTCTGCTTCTCTCCC	213836			
QY 2514	TGCCATCCCGAGGGAACCAATGCCATGG	2539			
Db 213837	TGCCATCCCGAGGGAACCAATGCCATGG	213862			
RESULT 11					
AC063964/c					
LOCUS					
DEFINITION	Homo sapiens chromosome 10 clone RP11-31303, WORKING DRAFT				
ACCESSION	AC063964				
VERSION	AC063964.3 GI:9887616				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Smith, D.R.				
JOURNAL	Genome Therapeutics Corporation Sequencing Center: Human Genome				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 175265)				
JOURNAL	Direct Submission				
AUTHORS	Submitted (22-APR-2000) Genome Therapeutics Corporation, 100 Beaver				
TITLE	Street, Waltham, MA 02453, USA				
JOURNAL	On Aug 24, 2000 this sequence version replaced gi:8439815.				
COMMENT	----- genome center				
	Center: Genome Therapeutics Corporation				
	Center code: GTC				
	Web site: http://www.genomecorp.com/				
	Contact: gtc-seqcenter@genomecorp.com				
	----- Project Information				
	----- Summary Statistics				
	Sequencing vector: N/A				
	Chemistry: Dye-terminator Big Dye; 100% of reads				
	Assembly program: Phrap; version 990315				
	Consensus quality: 153700 bases at least Q40				
	Consensus quality: 164398 bases at least Q30				
	Consensus quality: 167130 bases at least Q20				
	Insert size: 173464; sum-of-contigs				
	Quality coverage: 4.1x in Q20 bases; sum-of-contigs				

	* NOTE: This is a 'working draft' sequence. It currently				
	* consists of 20 contigs. The true order of the pieces				
	* is not known and their order in this sequence record is				
	* arbitrary. Gaps between the contigs are represented as				
	* runs of N, but the exact sizes of the gaps are unknown.				
	* This record will be updated with the finished sequence				
	* as soon as it is available and the accession number will				
	* be preserved.				
	1				
	* 1145: contig of 1145 bp in length				
	* 1146: gap of unknown length				
	* 1246: contig of 1169 bp in length				
	* 2415: gap of unknown length				
	* 2515: contig of 1350 bp in length				
	* 3864: gap of unknown length				
	* 3865: contig of 1114 bp in length				
	* 5078: gap of unknown length				
	* 5178: contig of 1469 bp in length				
	* 5179: gap of unknown length				
	* 6648: contig of 1179 bp in length				
	* 6748: gap of unknown length				
	* 7927: contig of 1386 bp in length				
	* 8027: gap of unknown length				
	* 9412: contig of 1386 bp in length				
	* 9413: gap of unknown length				

* 9513 10563: contig of 1051 bp in length
* 10564 10563: gap of unknown length
* 10664 11995: contig of 1332 bp in length
* 11996 12095: gap of unknown length
* 12096 13309: contig of 1214 bp in length
* 13310 13409: gap of unknown length
* 13410 15172: contig of 1763 bp in length
* 15173 15272: gap of unknown length
* 15273 17973: contig of 2701 bp in length
* 17974 18073: gap of unknown length
* 18074 21419: contig of 3346 bp in length
* 21420 21519: gap of unknown length
* 21520 24881: contig of 3362 bp in length
* 24882 24981: gap of unknown length
* 24982 31876: contig of 6895 bp in length
* 31877 31976: gap of unknown length
* 31977 49015: contig of 17039 bp in length
* 49016 49115: gap of unknown length
* 49116 66286: contig of 17171 bp in length
* 66287 66387: gap of unknown length
* 66387 86201: contig of 19815 bp in length
* 86202 86301: gap of unknown length
* 86302 132480: contig of 46179 bp in length
* 132481 132580: gap of unknown length
* 132581 175265: contig of 42685 bp in length.

FEATURES
source

1. .175265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-31303"
/clone_lib="RPC1-11"
1. .1145
/note="assembly_name:Contig1"
1246. .2414
/note="assembly_name:Contig11"
2515. .3864
/note="assembly_name:Contig12"
3965. .5078
/note="assembly_name:Contig15"
5179. .8647
/note="assembly_name:Contig18"
6748. .7926
/note="assembly_name:Contig20"
8027. .9412
/note="assembly_name:Contig23"
9513. .10563
/note="assembly_name:Contig25"
10664. .11995
/note="assembly_name:Contig27"
12096. .13309
/note="assembly_name:Contig28"
13410. .15172
/note="assembly_name:Contig29"
15273. .17973
/note="assembly_name:Contig30"
18074. .21419
/note="assembly_name:Contig31"
21520. .24881
/note="assembly_name:Contig32"
24982. .31876
/note="assembly_name:Contig33"
31977. .49015
/note="assembly_name:Contig34"
49116. .66286
/note="assembly_name:Contig35
clone_end:SP6"
66387. .86201
/note="assembly_name:Contig36"
86302. .132480
/note="assembly_name:Contig37
clone_end:T7"
132581. .175265

/note="assembly_name:Contig38"
BASE COUNT 53132 a 34451 c 33650 g 52102 t 1930 others
ORIGIN
Query Match 6.0%; Score 236; DB 2; Length 175265;
Best Local Similarity 100.0%; Pred. No. 2.3e-119;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1648 ACCACCAACCTGGCAGCTATTGAGCCACAGCCCTCCACCCACCTCAGAGCCCTCCATCT 1707
|||||
Db 54760 ACCACCAACCTGGCAGCTATTGAGCCACAGCCCTCCACCCACCTCAGAGCCCTCCATCT 54701
|||||
QY 1708 GTGGAACAAACCCCAACCCCAAACTCGAGGGGGTCTTGGTGAACCAACCAATGAGCCCGG 1767
|||||
Db 54700 GTGGAACAAACCCCAACCCCAAACTCGAGGGGGTCTTGGTGAACCAACCAATGAGCCCGG 54641
|||||
QY 1768 TCCAGCTCCAGGATTGGGCTTCGTGTGCACTTCAACTGCCTGAAGATGACAAGGAAGT 1827
|||||
Db 54640 TCCAGCTCCAGGATTGGGCTTCGTGTGCACTTCAACTGCCTGAAGATGACAAGGAAGT 54581
|||||
QY 1828 GAAGCATCTCCGAGGCTGGTGTGGTGACCAACAGACAGACAGCCGCGGATTCCTTT 1883
|||||
Db 54580 GAAGCATCTCCGAGGCTGGTGTGGTGACCAACAGACAGACAGCCGCGGATTCCTTT 54525
|||||
RESULT 12
AL442068/c 290625 bp DNA linear HTG 20-NOV-2001
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP11-331G8, *** SEQUENCING IN
PROGRESS ***, 43 unordered pieces.
ACCESSION AL442068
VERSION AL442068.16 GI:17043795
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Sims, S.
Direct Submission
Submitted (19-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 21, 2001 this sequence version replaced gi:14575284.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA331G8
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 277809 bases at least Q40
Consensus quality: 281712 bases at least Q30
Consensus quality: 284299 bases at least Q20
Insert size: 286425; sum-of-contigs
Quality coverage: 156064; 7.4% error; agarose-fp
Quality coverage: 5.79x in Q20 bases; sum-of-contigs Quality
coverage: 15.47x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 103320: contig of 103320 bp in length
* 103321 103420: gap of 100 bp

misc_feature /note="assembly_fragment:04166"
233117..236145
misc_feature /note="assembly_fragment:04570"
236246..239603
misc_feature /note="assembly_fragment:04709"
239710..241796
misc_feature /note="assembly_fragment:04939"
241897..243915

Query Match 5.7%; Score 227; DB 2; Length 290625;
Best Local Similarity 100.0%; Pred. No. 2.2e-114; Mismatches 0; Indels 0; Gaps 0;
Matches 227; Conservative 0;
QY 1648 ACCACAACTGCGACTATTGAGCCACAGCCCTCCACCCACCTCAGAGCGCTCCATCT 1707
Db 285418 ACCACAACTGCGACTATTGAGCCACAGCCCTCCACCCACCTCAGAGCGCTCCATCT 285359
QY 1708 GTGGAACAACCCCAACCCCAACCTCGAGGGGGTTCCTGGTGAACCAATGAGCCCCGG 1767
Db 285358 GTGGAACAACCCCAACCCCAACCTCGAGGGGGTTCCTGGTGAACCAATGAGCCCCGG 285299
QY 1768 TCACGCTCCAGGATTCGGCTTCGTGTCACCTCAACCTGCCTGAAGATGACAAGCAAT 1827
Db 285298 TCACGCTCCAGGATTCGGCTTCGTGTCACCTCAACCTGCCTGAAGATGACAAGCAAT 285239
QY 1828 GAAGCATCTCCGAGGCTGGTGTGGTGACCCAGACAGACAGGCGC 1874
Db 285238 GAAGCATCTCCGAGGCTGGTGTGGTGACCCAGACAGACAGGCGC 285192

RESULT 13
AC016395/c
LOCUS AC016395 166869 bp DNA linear PRI 14-MAY-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-153K11, complete sequence.
AC016395
VERSION AC016395.5 GI:20564406
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166869)
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166869)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 166869)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 166869)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 14, 2002 this sequence version replaced gi:9929646.
FEATURES
source 1..166869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-153K11"
/clone_lib="RPCL-11"
BASE COUNT 47408 a 35216 c 35336 g 48909 t
ORIGIN
Query Match 5.3%; Score 210; DB 9; Length 166869;

Best Local Similarity 100.0%; Pred. No. 7.1e-105;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3284 AGGTGAGTGGTTTACCGCCCGGAGCTGACATGGCTACTCAATGGCCACCTGTGCTAC 3343
Db 146665 AGGTGAGTGGTTTACCGCCCGGAGCTGACATGGCTACTCAATGGCCACCTGTGCTAC 146606
QY 3344 CAGATGCCTCCCAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTGACC 3403
Db 146605 CAGATGCCTCCCAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTGACC 146546
QY 3404 CACTCACTCAGCGCAGCGAGGACCTATAAGTGCATGCCACCAACCAACCGGCACA 3463
Db 146545 CACTCACTCAGCGCAGCGAGGACCTATAAGTGCATGCCACCAACCAACCGGCACA 146486
QY 3464 ATTCTTTTGTAGTCGGAGCTCTCTGTAGTAG 3493
Db 146485 ATTCTTTTGTAGTCGGAGCTCTCTGTAGTAG 146456
RESULT 14
AC114220/c
LOCUS AC114220 151002 bp DNA linear HTG 17-JUL-2002
DEFINITION Rattus norvegicus clone CH230-318023, *** SEQUENCING IN PROGRESS
***, 43 unordered pieces.
AC114220
VERSION AC114220.2 GI:21744220
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 151002)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,F.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.E.,
Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,B., Mawhiney,E., Mcleod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Roife,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooftari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 151002)
Worley, K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 151002)
Worley, K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:19224547.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOMH
Center clone name: CH230-318023
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 109622 bases at least Q40
Consensus quality: 113752 bases at least Q30
Consensus quality: 117081 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1050: contig of 1050 bp in length
* 1051 1150: gap of unknown length
* 1151 2234: contig of 1084 bp in length
* 2234 2235: gap of unknown length
* 2235 3515: contig of 1181 bp in length
* 3515 3615: gap of unknown length
* 3615 4929: contig of 1313 bp in length
* 4929 5028: gap of unknown length
* 5028 6455: contig of 1427 bp in length
* 6455 7769: contig of 1213 bp in length
* 7769 7869: gap of unknown length
* 7869 8881: contig of 1012 bp in length
* 8881 8981: gap of unknown length
* 8981 10397: contig of 1416 bp in length
* 10397 12153: gap of unknown length
* 12153 12254: contig of 1657 bp in length
* 12254 13855: contig of 1602 bp in length
* 13855 13956: gap of unknown length
* 13956 15470: contig of 1514 bp in length
* 15470 17115: gap of unknown length
* 17115 17215: contig of 1546 bp in length
* 17215 18930: gap of unknown length
* 18930 20071: contig of 1715 bp in length
* 20071 20171: gap of unknown length
* 20171 21498: contig of 1041 bp in length
* 21498 21599: gap of unknown length
* 21599 23009: contig of 1327 bp in length
* 23009 23009: contig of 1411 bp in length
* 23009 23009: contig of 1411 bp in length

23010 23109: gap of unknown length
23110 25272: contig of 2163 bp in length
25272 25373: gap of unknown length
25373 27230: contig of 1858 bp in length
27230 27331: gap of unknown length
27331 28961: contig of 1631 bp in length
28961 29061: gap of unknown length
29061 31889: contig of 2828 bp in length
31889 31890: gap of unknown length
31890 34607: contig of 2618 bp in length
34607 34707: gap of unknown length
34707 37331: contig of 2624 bp in length
37331 37431: gap of unknown length
37431 39898: contig of 2467 bp in length
39898 39998: gap of unknown length
39998 42294: contig of 2296 bp in length
42294 42394: gap of unknown length
42394 45156: contig of 2762 bp in length
45156 45256: gap of unknown length
45256 48350: contig of 3094 bp in length
48350 48451: gap of unknown length
48451 52367: contig of 3917 bp in length
52367 52468: gap of unknown length
52468 54222: contig of 1755 bp in length
54222 54323: gap of unknown length
54323 57294: gap of unknown length
57294 60595: contig of 3202 bp in length
60595 60696: gap of unknown length
60696 64285: contig of 3590 bp in length
64285 64385: gap of unknown length
64385 64386: contig of 4045 bp in length
64386 68431: gap of unknown length
68431 72515: contig of 3985 bp in length
72515 72615: gap of unknown length
72615 78326: contig of 5711 bp in length
78326 78427: gap of unknown length
78427 83087: contig of 4661 bp in length
83087 83187: gap of unknown length
83187 89628: contig of 6441 bp in length
89628 89729: gap of unknown length
89729 96247: contig of 6518 bp in length
96247 96347: gap of unknown length
96347 103841: contig of 7494 bp in length
103841 112875: contig of 8935 bp in length
112875 112975: gap of unknown length
112975 121627: contig of 8652 bp in length
121627 121727: gap of unknown length
121727 131119: contig of 9392 bp in length
131119 131220: gap of unknown length
131220 139984: contig of 8765 bp in length
139984 140084: gap of unknown length
140084 151002: contig of 10918 bp in length.
151002 151002: contig of 10918 bp in length.
Location/Qualifiers
1. 151002
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-318023"
BASE COUNT 41237 a 32621 c 32087 g 40808 t 4249 others
ORIGIN
Query Match 1.0%; Score 41; DB 2; Length 151002;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3331 CAACCTGTGTACAGATGCTCCACAGATGCTGTCAG 3371
|||||
Db 126983 CAACCTGTGTACAGATGCTCCACAGATGCTGTCAG 126943
|||||
RESULT 15
AC114220

LOCUS AC114220 151002 bp DNA linear HTG 17-JUL-2002
 DEFINITION Rattus norvegicus clone CH230-318023, *** SEQUENCING IN PROGRESS
 AC114220 *** 43 unordered pieces.
 VERSION AC114220.2 GI:21744220
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 151002)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaraturge, H.C., Are, J.R., Ayale, M., Banks, T.,
 Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Hollaway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, P., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okwuonu, G.,
 Oraguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 151002)
 Worley, K.C.
 Direct Submission
 Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 151002)
 Worley, K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 13, 2002 this sequence version replaced gi:19224547.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

Center project name: GQWH
 Center clone name: CH230-318023
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 109622 bases at least Q40
 Consensus quality: 113752 bases at least Q30
 Consensus quality: 117081 bases at least Q20

 NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 43 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1050: contig of 1050 bp in length
 1051 1150: gap of unknown length
 1151 2234: contig of 1084 bp in length
 2235 2335: gap of unknown length
 2335 3515: contig of 1181 bp in length
 3516 3615: gap of unknown length
 3616 4929: contig of 1313 bp in length
 4929 5028: gap of unknown length
 5029 6455: contig of 1427 bp in length
 6456 6555: gap of unknown length
 6556 7769: contig of 1213 bp in length
 7769 7868: gap of unknown length
 7869 8880: contig of 1012 bp in length
 8881 8980: gap of unknown length
 8981 10396: contig of 1416 bp in length
 10397 10497: gap of unknown length
 10497 12154: contig of 1657 bp in length
 12154 12254: gap of unknown length
 12254 13855: contig of 1602 bp in length
 13856 13955: gap of unknown length
 13956 15469: contig of 1514 bp in length
 15470 15569: gap of unknown length
 15570 17115: contig of 1546 bp in length
 17116 17215: gap of unknown length
 17216 18930: contig of 1715 bp in length
 18931 19030: gap of unknown length
 19031 20071: contig of 1041 bp in length
 20072 20171: gap of unknown length
 20172 21498: contig of 1327 bp in length
 21499 21598: gap of unknown length
 21599 23109: contig of 1411 bp in length
 23110 23109: gap of unknown length
 23110 25272: contig of 2163 bp in length
 25273 25372: gap of unknown length
 25373 27230: contig of 1858 bp in length
 27231 27330: gap of unknown length
 27331 28961: contig of 1631 bp in length
 28962 29061: gap of unknown length
 29062 31889: contig of 2828 bp in length
 31890 31889: gap of unknown length
 31890 34607: contig of 2618 bp in length
 34608 34707: gap of unknown length
 34708 37331: contig of 2624 bp in length
 37332 37431: gap of unknown length
 37432 39898: contig of 2467 bp in length
 39899 39998: gap of unknown length
 42294: contig of 2296 bp in length
 42295 42394: gap of unknown length
 42395 45156: contig of 2762 bp in length
 45157 45256: gap of unknown length
 45257 48350: contig of 3094 bp in length
 48351 48450: gap of unknown length
 48451 52367: contig of 3917 bp in length

```
* 52368 52467: gap of unknown length
* 52468 54222: contig of 1755 bp in length
* 54323 54323: gap of unknown length
* 54323 57293: contig of 2971 bp in length
* 57294 57393: gap of unknown length
* 57394 60595: contig of 3202 bp in length
* 60596 60695: gap of unknown length
* 60696 64285: contig of 3590 bp in length
* 64286 64385: gap of unknown length
* 64386 68430: contig of 4045 bp in length
* 68431 68530: gap of unknown length
* 68531 72515: contig of 3985 bp in length
* 72516 72616: gap of unknown length
* 72616 78326: contig of 5711 bp in length
* 78327 78426: gap of unknown length
* 78427 83087: contig of 4661 bp in length
* 83088 83187: gap of unknown length
* 83188 89628: contig of 6441 bp in length
* 89629 89728: gap of unknown length
* 89729 96246: contig of 6518 bp in length
* 96247 96346: gap of unknown length
* 96347 103840: contig of 7494 bp in length
* 103841 103940: gap of unknown length
* 103941 112875: contig of 8935 bp in length
* 112876 112975: gap of unknown length
* 112976 121627: contig of 8652 bp in length
* 121628 121727: gap of unknown length
* 121728 131119: contig of 9392 bp in length
* 131120 131219: gap of unknown length
* 131220 139984: contig of 8765 bp in length
* 139985 140084: gap of unknown length
* 140085 151002: contig of 10918 bp in length.
```

```
FEATURES             Location/Qualifiers
    source             1..151002
                        /organism="Rattus norvegicus"
                        /db_xref="taxon:10116"
                        /clone="CH230-318023"
```

```
BASE COUNT   41237 a 32621 c 32087 g 40808 t 4249 others
ORIGIN
```

```
Query Match          1.0%; Score 38; DB 2; Length 151002;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2890 TTCACCTGCAAAATTGTTGGGATACCTGTTCCAAAGGT 2927
      |||||
Db 41465 TTCACCTGCAAAATTGTTGGGATACCTGTTCCAAAGGT 41502
```

Search completed: November 30, 2002, 16:55:12
Job time : 9637 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 13:41:14 ; Search time 516 Seconds
(without alignments)
17295.871 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 3963
Sequence: 1 atgcaagacgacgacataga.....tggagagtgatgaactttaa 3963

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :					N_Geneseq_l010002:*					
					1:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*				
					2:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*				
					3:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*				
					4:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*				
					5:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*				
					6:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*				
					7:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*				
					8:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*				
					9:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*				
					10:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*				
					11:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*				
					12:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*				
					13:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*				
					14:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*				
					15:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*				
					16:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*				
					17:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*				
					18:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*				
					19:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*				
					20:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*				
					21:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*				
					22:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*				
					23:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*				
					24:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*				
										SUMMARIES
Result No.	Score	Query Match	Length	DB	ID	Description				
1	1483	37.4	2127	22	AAH13889	Human cDNA sequenc				
2	1056	26.6	1228	23	AAS74798	DNA encoding novel				
3	787	19.9	2059	22	ABA06552	Human cDNA SEQ ID				
4	685	17.3	2232	22	AAS28777	Human immunoglobul				
5	438	11.1	855	22	AAH04212	Human cDNA clone (
6	348	8.8	702	22	ABA06731	Human cDNA SEQ ID				
7	348	8.8	702	22	AAS28841	Human immunoglobul				
8	318	8.0	392	21	AACT5084	Human ORFX ORF639				
9	99	2.5	200	24	ABN25334	Human ORFX polynuc				
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.										

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

667	17	0.4	2132	22	AA013237	Human BCL-X-like p	c 740	17	0.4	2784	23	AA572946	DNA encoding novel
668	17	0.4	2133	22	AA526108	Human cDNA encoding	741	17	0.4	2784	23	AA577185	DNA encoding novel
c 669	17	0.4	2136	22	AA526530	Human cDNA encoding	c 742	17	0.4	2797	22	AA844603	Human zF3 protein
c 670	17	0.4	2139	17	AA741704	Lymphocyte specific	743	17	0.4	2808	16	AA705764	Human E-cadherin c
671	17	0.4	2141	22	AA945409	Human full-length	744	17	0.4	2808	24	ABN89392	Human E-cadherin e
672	17	0.4	2161	23	ABL20939	Drosophila melanog	745	17	0.4	2815	15	AA065487	Sequence of human
c 673	17	0.4	2162	22	AA016756	Human novel protei	746	17	0.4	2843	24	ABK36030	cdna sequence #421
674	17	0.4	2165	21	AA295313	Human mitochondria	c 747	17	0.4	2849	20	AA800058	Human PRO353 nucle
675	17	0.4	2165	24	ABK84712	Human cDNA differe	c 748	17	0.4	2849	21	AA449566	Human PRO353 cDNA
676	17	0.4	2165	24	ABN97328	Gene #3826 used to	c 749	17	0.4	2849	21	AA449566	Human cDNA sequenc
677	17	0.4	2179	23	ABL09929	Drosophila melanog	c 750	17	0.4	2849	22	AA87013	Nucleotide sequenc
c 678	17	0.4	2181	17	AA112500	Non-infective tick	c 751	17	0.4	2859	23	ABL04595	Drosophila melanog
c 679	17	0.4	2181	19	AAV59451	Plasmid construct	c 752	17	0.4	2868	21	AA958825	Human metalloprote
680	17	0.4	2206	24	AAK24325	DNA encoding human	753	17	0.4	2898	24	ABT99437	Mouse ischaemic co
c 681	17	0.4	2214	22	AAH68404	C glutamicum codin	c 754	17	0.4	2948	23	ABL15091	Drosophila melanog
c 682	17	0.4	2249	23	ABV24622	Human prostate exp	755	17	0.4	2948	24	AA145975	Human macro protei
c 683	17	0.4	2253	23	ABL29233	Drosophila melanog	c 756	17	0.4	2952	22	AA16759	Human novel protei
684	17	0.4	2259	22	AA522473	Human cDNA encoding	c 757	17	0.4	2976	22	AA580195	DNA encoding novel
685	17	0.4	2260	20	AA241314	Human normal ovari	758	17	0.4	2995	22	AA158266	Human polynucleoti
c 686	17	0.4	2270	21	AAK75892	Human ORFX ORF1447	c 759	17	0.4	3035	23	ABL05379	Drosophila melanog
c 687	17	0.4	2276	20	AA355550	DNA encoding a F-b	760	17	0.4	3105	22	AAH52200	Human AFP protein
c 688	17	0.4	2276	22	AA014885	Human full length	761	17	0.4	3108	24	ABA04428	Human intracellular
689	17	0.4	2277	22	AA844459	Interleukin-1 rece	762	17	0.4	3143	24	AA337608	Human p1045 prote
690	17	0.4	2288	21	AA561602	DNA encoding an in	c 763	17	0.4	3185	23	AA591627	DNA encoding novel
691	17	0.4	2293	22	AA157800	Human polynucleoti	c 764	17	0.4	3196	22	AAH98392	Human EST-derived
692	17	0.4	2295	22	AA522788	Human cDNA encoding	765	17	0.4	3198	23	ABL15738	Drosophila melanog
c 693	17	0.4	2332	22	AAH17008	Human cDNA sequenc	766	17	0.4	3199	23	ABL21330	Drosophila melanog
c 694	17	0.4	2337	22	AA71634	Corynebacterium gl	767	17	0.4	3221	23	ABL06510	Drosophila melanog
c 695	17	0.4	2343	22	AA945524	Human full-length	768	17	0.4	3232	22	AAH54722	S. epidermidis gen
696	17	0.4	2362	20	AA200357	Nucleotide sequenc	c 769	17	0.4	3276	19	AAV59791	Human secreted pro
697	17	0.4	2362	22	AA8F9019	Human FATP2 coding	770	17	0.4	3376	20	AAK81414	Intron sequence of
c 698	17	0.4	2368	17	AA742128	Neisseria gonorrhoe	c 771	17	0.4	3386	24	ABK84112	Human cDNA differe
c 699	17	0.4	2381	22	AA010202	N. gonorrhoeae or N	-c 772	17	0.4	3386	24	ABK64551	Human benign prost
c 700	17	0.4	2418	9	AA81162	Encodes Western su	773	17	0.4	3389	23	ABV22266	Human prostate exp
701	17	0.4	2430	22	ABA15899	Human nervous syst	774	17	0.4	3389	23	ABV28103	Human prostate exp
702	17	0.4	2430	22	ABA16361	Human soluble guan	c 775	17	0.4	3413	22	AAK66240	Human immune/haema
703	17	0.4	2443	21	AA288939	DNA encoding novel	c 776	17	0.4	3460	19	AAV36957	Nucleotide sequenc
704	17	0.4	2456	23	AA574097	DNA encoding novel	c 777	17	0.4	3460	21	AA234989	Human cell surface
c 705	17	0.4	2461	22	AA575923	Human cell cycle a	c 778	17	0.4	3460	24	ABK84482	Human cDNA differe
706	17	0.4	2471	22	AAK94363	Human full-length	779	17	0.4	3478	23	ABL15826	Drosophila melanog
707	17	0.4	2490	23	AA586902	DNA encoding novel	c 780	17	0.4	3509	22	AAH15777	Human cDNA sequenc
708	17	0.4	2509	23	ABL20287	Drosophila melanog	c 781	17	0.4	3523	23	ABL09879	Drosophila melanog
709	17	0.4	2512	23	AA567345	DNA encoding novel	c 782	17	0.4	3549	23	ABL10635	Drosophila melanog
c 710	17	0.4	2529	21	AAAC42940	Arabidopsis thalia	c 783	17	0.4	3597	23	ABL09813	Drosophila melanog
c 711	17	0.4	2532	24	ABK11656	DNA encoding human	c 784	17	0.4	3657	23	ABV25797	Human prostate exp
c 712	17	0.4	2553	24	ABK12900	Human protease PRT	785	17	0.4	3670	22	AA160052	Human polynucleoti
c 713	17	0.4	2553	24	ABK31792	DNA encoding novel	c 786	17	0.4	3703	22	AA87636	Human RecQ5 beta D
c 714	17	0.4	2590	21	AA236239	cdna encoding a bo	c 787	17	0.4	3711	23	ABL26102	Drosophila melanog
c 715	17	0.4	2616	14	AAQ37966	Sequence encoding	c 788	17	0.4	3776	24	ABQ54759	Human ovarian anti
716	17	0.4	2617	22	AA526104	Human cDNA encoding	c 789	17	0.4	3807	23	ABV21514	Human prostate exp
c 717	17	0.4	2627	24	ABL64775	Lung cancer relate	c 790	17	0.4	3807	23	ABV22724	Human prostate exp
c 718	17	0.4	2627	24	ABL65409	Lung cancer relate	c 791	17	0.4	3807	23	ABV24950	Human prostate exp
c 719	17	0.4	2627	24	ABL66513	Lung cancer relate	c 792	17	0.4	3807	23	ABV24965	Human prostate exp
c 720	17	0.4	2627	24	ABL68130	Ovary cancer relat	c 793	17	0.4	3807	23	ABV25855	Human prostate exp
c 721	17	0.4	2627	24	ABL69119	Kidney cancer rela	c 794	17	0.4	3807	23	ABV27332	Human prostate exp
c 722	17	0.4	2631	24	AA596334	Arabidopsis cDNA e	c 795	17	0.4	3807	23	ABV28549	Human prostate exp
c 723	17	0.4	2638	11	AAQ03741	Human myb related	c 796	17	0.4	3900	23	AA564300	DNA encoding novel
c 724	17	0.4	2638	19	AAV20468	Human B-myb oncoge	c 797	17	0.4	4043	23	ABL22462	Drosophila melanog
725	17	0.4	2640	23	ABL15830	Drosophila melanog	c 798	17	0.4	4076	23	ABL02152	Drosophila melanog
726	17	0.4	2641	22	AA161175	Human polynucleoti	c 799	17	0.4	4082	22	AAK79954	Human immune/haema
727	17	0.4	2643	20	AA331982	Sulfated fucose-co	c 800	17	0.4	4082	22	AAK82476	Human immune/haema
728	17	0.4	2644	24	ABN59851	Novel human coding	c 801	17	0.4	4085	22	AAK79956	Human immune/haema
729	17	0.4	2650	23	ABL29347	Drosophila melanog	c 802	17	0.4	4085	22	AAK82478	Human immune/haema
c 730	17	0.4	2651	19	AAV52349	Streptococcus pneu	c 803	17	0.4	4086	22	AAH49373	Human metalloprote
731	17	0.4	2654	23	ABL19179	Drosophila melanog	c 804	17	0.4	4086	22	AAK79955	Human immune/haema
c 732	17	0.4	2656	24	AA594947	Human DNA sequenc	c 805	17	0.4	4086	22	AAK82477	Human immune/haema
c 733	17	0.4	2661	22	AA159389	Human polynucleoti	c 806	17	0.4	4123	24	ABK50067	DNA encoding mouse
c 734	17	0.4	2735	23	AA585641	DNA encoding novel	c 807	17	0.4	4148	21	AAA46556	cDNA sequence enco
c 735	17	0.4	2748	23	ABL09795	Drosophila melanog	c 808	17	0.4	4148	24	ABK72679	DNA encoding enco
736	17	0.4	2749	24	ABL34423	Human immune syste	c 809	17	0.4	4260	23	AA574826	DNA encoding novel
c 737	17	0.4	2754	23	AA568677	DNA encoding novel	c 810	17	0.4	4270	24	ABN95930	Gene #2428 used to
738	17	0.4	2769	20	AA224813	Human adipocyte-sp	811	17	0.4	4270	24	ABL64082	Breast cancer rela
c 739	17	0.4	2769	21	AAA64114	DNA encoding a hum	812	17	0.4	4283	22	AA525994	Human cDNA encodin

Db 651 TGTTCTGAACCCAGTCCAAAAAAGTATTTTAAATTAAGGCTGCCGACTTCATTGAAGAG 710
QY 481 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCGTGCCTGCCAAAAC 540
Db 711 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCGTGCCTGCCAAAAC 770
QY 541 CACAAGAGTAACTGGAATCTCAAAACAAAGTTATGCGAGAAACAGCTCCAGTTTCTCA 600
Db 771 CACAAGAGTAACTGGAATCTCAAAACAAAGTTATGCGAGAAACAGCTCCAGTTTCTCA 830
QY 601 GATCTGTCAGAAAGCAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCAG 660
Db 831 GATCTGTCAGAAAGCAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCAG 890
QY 661 GATAATGAAGTGAATCACGCCCTGGAAACAGCAGGAGCCAAAGAGGCTGAAGCGGAGCAG 720
Db 891 GATAATGAAGTGAATCACGCCCTGGAAACAGCAGGAGCCAAAGAGGCTGAAGCGGAGCAG 950
QY 721 GCTGCCAGTGAGCGGCTGGTGGAGACATACACCAAGGCTTCCCTTTTCATCTCTGTAC 780
Db 951 GCTGCCAGTGAGCGGCTGGTGGAGACATACACCAAGGCTTCCCTTTTCATCTCTGTAC 1010
QY 781 TATGAAGAACCTCTGGGCAACCTCCCGGTTCACTCAAAAGTTAGCGGACGAGAGTT 840
Db 1011 TATGAAGAACCTCTGGGCAACCTCCCGGTTCACTCAAAAGTTAGCGGACGAGAGTT 1070
QY 841 CCAGAAGGAACCTCGACTACAGTTGGATTGCATAGTGTAGGAATTCACACCTCAAGTA 900
Db 1071 CCAGAAGGAACCTCGACTACAGTTGGATTGCATAGTGTAGGAATTCACACCTCAAGTA 1130
QY 901 AGGTGCTACTGTGAAGCAGAGCTTGAAAATTTCCCCAGATATTCACATCGTCCAGGCA 960
Db 1131 AGGTGCTACTGTGAAGCAGAGCTTGAAAATTTCCCCAGATATTCACATCGTCCAGGCA 1190
QY 961 GGAAATCTGCATCACTGACCTGCGGAGCCTTTGAGAGGACACAGACGCTATTCC 1020
Db 1191 GGAAATCTGCATCACTGACCTGCGGAGCCTTTGAGAGGACACAGACGCTATTCC 1250
QY 1021 TGCTTTGCTCTTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1080
Db 1251 TGCTTTGCTCTTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1310
QY 1081 GTTCTTCTTCTGACTCAGAGGCGCCCTTAACAAGGAAGAGATGAATCGAATCCAGAAG 1140
Db 1311 GTTCTTCTTCTGACTCAGAGGCGCCCTTAACAAGGAAGAGATGAATCGAATCCAGAAG 1370
QY 1141 CCNAATGAGGTGTCATCTCTCCACTACCTCTGCAAGTATTCCTCCAGCAGTACCCCAA 1200
Db 1371 CCNAATGAGGTGTCATCTCTCCACTACCTCTGCAAGTATTCCTCCAGCAGTACCCCAA 1430
QY 1201 GCCCAGCATTTGGTGCCCAACCTCTGTGGCAACCATCCAGCAGTGTGAGAGCCCCACC 1260
Db 1431 GCCCAGCATTTGGTGCCCAACCTCTGTGGCAACCATCCAGCAGTGTGAGAGCCCCACC 1490
QY 1261 AATTACTTCAGGGATTGGATGGAACCTATCATTTGCAGCTCCTGTGTTTACAAAGATG 1320
Db 1491 AATTACTTCAGGGATTGGATGGAACCTATCATTTGCAGCTCCTGTGTTTACAAAGATG 1550
QY 1321 CTACAAAATTTGACCTTCTGAGGTCAGCTGGTGTCTTTGAATGCAGAGTAAAGGA 1380
Db 1551 CTACAAAATTTGACCTTCTGAGGTCAGCTGGTGTCTTTGAATGCAGAGTAAAGGA 1610
QY 1381 GCTCCATCTCTAAGGTTGAGTGGTATAGAGAAGGACTTTAATAGAAGATTTCTCCAGAT 1440
Db 1611 GCTCCATCTCTAAGGTTGAGTGGTATAGAGAAGGACTTTAATAGAAGATTTCTCCAGAT 1670
QY 1441 TTTAGGATTTTACAGAAAAACCTTCGATCCATCGCAGAGCCAG 1483
Db 1671 TTTAGGATTTTACAGAAAAACCTTCGATCCATCGCAGAGCCAG 1713

RESULT 2
AAS74798

ID AAS74798 standard; cDNA; 1228 BP.
XX AC AAS74798;
XX XX
DT 13-FEB-2002 (first entry)
XX XX
DE DNA encoding novel human diagnostic protein #10602.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX XX
OS Homo sapiens.
PN WO200175067-A2.
XX XX
PD 11-OCT-2001.
XX XX
PF 30-MAR-2001; 2001WO-US08631.
XX XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG10611.
XX XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX XX
PS Claim 1: SEQ ID No 10602; 103pp; English.
XX XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94584 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 1228 BP; 323 A; 319 C; 314 G; 272 T; 0 other;

Query Match 26.6%; Score 1056; DB 23; Length 1228;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1236; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2737 ATGAATGAATAGAGTTTCCTTGGAAACGACTCTCTTGTATGAATCAGATGATGAATT 2796
Db 1 ATGAATGAATAGAGTTTCCTTGGAAACGACTCTCTTGTATGAATCAGATGATGAATT 60
QY 2797 CAACATGATGAGATCCCGGCGCAAGTGTATTCCTCCCATCTTGACAGAGACTCAAG 2856
Db 61 CAACATGATGAGATCCCGGCGCAAGTGTATTCCTCCCATCTTGACAGAGACTCAAG 120
QY 2857 CACTTCGGGTCACAGAAGGCTCTCCAGTCACATTCACCTGCAAAATTTGTTGGGATACCT 2916

Db 121 CACTTCGGGTACAGAGAGCTCTCCAGTTACATTCACCTCAAAATTTGTTGGATACCT 180
|||||
Qy 2917 GTTCCAAAGGTTTACTGGTTCAAGATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGC 2976
|||||
Db 181 GTTCCAAAGGTTTACTGGTTCAAGATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGC 240
|||||
Qy 2977 AAAATGAGCGAGAGAGATGGGACATGCTCTCTGCACATTTGAATCCACTACCAGTGAT 3036
|||||
Db 241 AAAATGAGCGAGAGAGATGGGACATGCTCTCTGCACATTTGAATCCACTACCAGTGAT 300
|||||
Qy 3037 GACGATGGCAACTACACCATCATGGCAGCCAAACCCCGAGGGGAGAATCAGCTGTTCTGGC 3096
|||||
Db 301 GACGATGGCAACTACACCATCATGGCAGCCAAACCCCGAGGGGAGAATCAGCTGTTCTGGC 360
|||||
Qy 3097 CACTTGATGTTACAAAGTTTGGCCATTTCGAGTCGGCTACCTCTGCTGGTCACTCTCAC 3156
|||||
Db 361 CACTTGATGTTACAAAGTTTGGCCATTTCGAGTCGGCTACCTCTGCTGGTCACTCTCAC 420
|||||
Qy 3157 AGGGGAAGATCCCGAGTGCAGAAAGAGACAAAGAGCCCTACAGAGCGCTTTTCCGA 3216
|||||
Db 421 AGGGGAAGATCCCGAGTGCAGAAAGAGACAAAGAGCCCTACAGAGCGCTTTTCCGA 480
|||||
Qy 3217 CCACATTTCCCTGCAGGCTCCTGGGATATGTTAGCTCATGAGGGGCGCCTCTGTCGGCTG 3276
|||||
Db 481 CCACATTTCCCTGCAGGCTCCTGGGATATGTTAGCTCATGAGGGGCGCCTCTGTCGGCTG 540
|||||
Qy 3277 GACTGTAAGTGAAGTGTATACCCGCCCGCGAGC-TGACATGGCTACTCAATGGCCAAAC 3335
|||||
Db 541 GACTGTAAGTGAAGTGTATACCCGCCCGCGAGCATGACATGGCTACTCAATGGCCAAAC 600
|||||
Qy 3336 TGTGCTACCAAGATGCTTCCCAAGATGCTGTGTCAGGAGACCGAGTCCACTCTCTGCT 3395
|||||
Db 601 TGTGCTACCAAGATGCTTCCCAAGATGCTGTGTCAGGAGACCGAGTCCACTCTCTGCT 660
|||||
Qy 3396 CATTGACCCACTACTCAGCGCGACGAGGACCTATAAGTGCATCGCTACCAACAAAC 3455
|||||
Db 661 CATTGACCCACTACTCAGCGCGACGAGGACCTATAAGTGCATCGCTACCAACAAAC 720
|||||
Qy 3456 CGGCGCAATCTTTTAGTCTGGAGCTCTCTGTAGTAGCCAAAGAGGTGAAGAAGCAC 3515
|||||
Db 721 CGGCGCAATCTTTTAGTCTGGAGCTCTCTGTAGTAGCCAAAGAGGTGAAGAAGCAC 780
|||||
Qy 3516 TGTGATCCTGGAGAACTACAGAACTCGGTTGTCCGAGGCGCCCGGTGAGACTGGA 3575
|||||
Db 781 TGTGATCCTGGAGAACTACAGAACTCGGTTGTCCGAGGCGCCCGGTGAGACTGGA 840
|||||
Qy 3576 GTGCCCGGTGATAGGATGCCCCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCAT 3635
|||||
Db 841 GTGCCCGGTGATAGGATGCCCCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCAT 900
|||||
Qy 3636 CCCTTGACACAGAGAGGATCAGTATGCACAGGACACACAGGGTATGCGTCCCTTCT 3695
|||||
Db 901 CCCTTGACACAGAGAGGATCAGTATGCACAGGACACACAGGGTATGCGTCCCTTCT 960
|||||
Qy 3696 CATTACCCAGCCAAAGAAATCAGACGCTGGATGTTACACGTTGTGACCCCAAGATGAAGC 3755
|||||
Db 961 CATTACCCAGCCAAAGAAATCAGACGCTGGATGTTACACGTTGTGACCCCAAGATGAAGC 1020
|||||
Qy 3756 CGGCATCGTGTGTCGACTGCCAGGCTGGATATACGCTCAGTGGCAACCATCAGATCCC 3815
|||||
Db 1021 CGGCATCGTGTGTCGACTGCCAGGCTGGATATACGCTCAGTGGCAACCATCAGATCCC 1080
|||||
Qy 3816 ACCGCCATGCTGTCCGGCCCGAGTGCAGTCGCTACGGATCTCTCAGAGTAAGGACT 3875
|||||
Db 1081 ACCGCCATGCTGTCCGGCCCGAGTGCAGTCGCTACGGATCTCTCAGAGTAAGGACT 1140
|||||
Qy 3876 TGACATATTTTCTGCCCTTTTCCCTCCATGGAAGACCATGCTGTTATTCATGCTCTTCTCG 3935
|||||
Db 1141 TGACATATTTTCTGCCCTTTTCCCTCCATGGAAGACCATGCTGTTATTCATGCTCTTCTCG 1200
|||||
Qy 3936 GAGTGTAGTGAGAGTGAATGAACCTTTAA 3963
|||||

Db 1201 GAGTGTAGTGGAGAGTGATGAACCTTAA 1228
RESULT 3
ABA06552
ID ABA06552 standard; cDNA; 2059 BP.
XX
AC ABA06552;
XX
XX 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 218.
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation; ss.
OS Homo sapiens.
XX
XX WO200154474-A2.
XX
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01349.
XX
XX 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225288P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226686P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.


```
Qy 3657 CAGTATGCACACAGACACAAACAGGATATGCTGCTTCTCATTCAGCCAGCCAAAGAATC 3716
|||||
Db 489 CAGTATGCACACAGACACAAACAGGATATGCTGCTTCTCATTCAGCCAGCCAAAGAATC 548
|||||
Qy 3717 AGACGCTGGATGGTACACGTTGTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGCACTGC 3776
|||||
Db 549 AGACGCTGGATGGTACACGTTGTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGCACTGC 608
|||||
Qy 3777 CAGGCTGGATATACGCTCAGTGGCACCACATCAGATCCACCCGCCATGTGTGCCGGCC 3836
|||||
Db 609 CAGGCTGGATATACGCTCAGTGGCACCACATCAGATCCACCCGCCATGTGTGCCGGCC 668
|||||
Qy 3837 CAGTGGCAGTGGTACCGATCTCTCACCAGTAAAGGACTTGACATATTTCTGCTTTTC 3896
|||||
Db 669 CAGTGGCAGTGGTACCGATCTCTCACCAGTAAAGGACTTGACATATTTCTGCTTTTC 728
|||||
Qy 3897 CTCATGGAAGCAGATGATGATTCATGCTCTTCGGAGTGTAGTGAGAGTGATGA 3956
|||||
Db 729 CTCATGGAAGCAGATGATGATTCATGCTCTTCGGAGTGTAGTGAGAGTGATGA 788
|||||
Qy 3957 ACTTTAA 3963
|||||
Db 789 ACTTTAA 795
|||||

RESULT 4
AAS28777
ID AAS28777 standard; cDNA; 2232 BP.
XX AC AAS28777;
XX DT 07-NOV-2001 (first entry)
XX DE Human immunoglobulin encoding cDNA SEQ ID No 23.
XX KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX OS Homo sapiens.
XX PN WO200155315-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01326.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
```

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-457725/49.
DR P-PSDB; AUI17989.
XX

PT Isolated novel immunoglobulin polypeptide for monitoring the presence
of and progression of diseases and for diagnosis -

XX Claim 1; SEQ ID No 23; 551pp; English.

XX Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the
immunoglobulin polypeptides of the invention, and primers for the
polynucleotides. The polynucleotides and polypeptides can be used to
diagnose a pathological condition or a susceptibility to a pathological
condition in a subject by determining the presence or absence of a
mutation in a DNA sequence or determining the presence or amount of
expression of the protein. Alternatively the identification of a binding
partner to a sequence allows determination of changes in protein
activity. The sequences can be used as research tools for receptors or
other signal transduction pathway proteins that interact with the
polypeptides of the invention and can be used to treat, prevent or
diagnose various types of disorders such as neurological disorders,
cardiovascular disorders, gastrointestinal disorders, reproductive
disorders, immune system disorders, renal disorders, muscular disorders,
pulmonary disorders, proliferative disorders and cancer.
XX Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX

XX Sequence 2232 BP; 617 A; 508 C; 528 G; 569 T; 10 other;

Query Match 17.3%; Score 685; DB 22; Length 2232;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 785; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3177 AGAAAGAGACAAAGAGCCCTACAGGAACGCTTTTCCGACACACATTTCTCTGAGGCTCC 3236
DB 2 AGAAAGAGACAAAGAGCCCTACAGGAACGCTTTTCCGACACACATTTCTCTGAGGCTCC 61
QY 3237 TGGGGATATGGTAGCTCATGAGGGCGCCTCTCTGGCTGGACTGTAAGGTGAGTGGTTT 3296
DB 62 TGGGGATATGGTAGCTCATGAGGGCGCCTCTCTGGCTGGACTGTAAGGTGAGTGGTTT 121
QY 3297 ACCGCCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTCTACAGATGCCTCCCA 3356
DB 122 ACCGCCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTCTACAGATGCCTCCCA 181
QY 3357 CAAGATGCTGGTCAGGGAGACCGGAGTCCACTCTCTGTCTCAATGACCCACTCACTCAGCG 3416
DB 182 CAAGATGCTGGTCAGGGAGACCGGAGTCCACTCTCTGTCTCAATGACCCACTCACTCAGCG 241
QY 3417 CGACGCGGAGCTATAAGTGCATCGCTACCAACAAACCGGCGAGAAATCTTTTAGTCT 3476
DB 242 CGACGCGGAGCTATAAGTGCATCGCTACCAACAAACCGGCGAGAAATCTTTTAGTCT 301
QY 3477 GGAGCTCTCTGTAGTCCAAAGAGGTGAAGAACACCTGTGTATCCTGGAGAACTACA 3536
DB 302 GGAGCTCTCTGTAGTCCAAAGAGGTGAAGAACACCTGTGTATCCTGGAGAACTACA 361
QY 3537 GAACTGCGGTGTCCCGAAGGCCACCCCGTGAGACTGGAGTGCCTGATAGGCGATGCC 3596
DB 362 GAACTGCGGTGTCCCGAAGGCCACCCCGTGAGACTGGAGTGCCTGATAGGCGATGCC 421
QY 3597 CCCACCTGTCTTCTACTGGAAGAACACATGAGACCATCCCTTGACACGAGAGAGAT 3656
DB 422 CCCACCTGTCTTCTACTGGAAGAACACATGAGACCATCCCTTGACACGAGAGAGAT 481
QY 3657 CAGTATGCACAGGACACACAGGGTATGCTGCTCTTCTCATTACGACCCCAAGAAATC 3716
DB 482 CAGTATGCACAGGACACACAGGGTATGCTGCTCTTCTCATTACGACCCCAAGAAATC 541
QY 3717 AGACGCTGGATGGTACACGCTTGTACGCCAAGAAATGAAGCGGCGATCGTGTGCGACTGC 3776
DB 542 AGACGCTGGATGGTACACGCTTGTACGCCAAGAAATGAAGCGGCGATCGTGTGCGACTGC 601
QY 3777 CAGGCTGGATATATACGCTCAGTGGCACCACATCAGATCCACCCCGCATGCTCTCGCGCC 3836
DB 602 CAGGCTGGATATATACGCTCAGTGGCACCACATCAGATCCACCCCGCATGCTCTCGCGCC 661
QY 3837 CAGTGGCAGTGGTACGCGATCTCTCACCAGTAAGGACTTGACATATTTTCTGCTTTTC 3896
DB 662 CAGTGGCAGTGGTACGCGATCTCTCACCAGTAAGGACTTGACATATTTTCTGCTTTTC 721
QY 3897 CTCCATGGAAGACGATGGTGTATTCTCTCGAGTGTAGTGAGAGATGATGA 3956
DB 722 CTCCATGGAAGACGATGGTGTATTCTCTCGAGTGTAGTGAGAGATGATGA 781
QY 3957 ACTTTAA 3963
DB 782 ACTTTAA 788

RESULT 5

AAH04212

ID AAH04212 standard; cDNA; 855 BP.

XX

AC AAH04212;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA clone (5'-primer) SEQ ID NO:1047.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

```
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX DR
XX PR
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS
XX PS Claim 1; SEQ ID 1047; 2537pp + CD ROM; English.
XX CC
XX CC The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ
XX SQ Sequence 855 BP; 257 A; 214 C; 176 G; 205 T; 3 other;
Query Match 11.1%; Score 438; DB 22; Length 855;
Best Local Similarity 99.8%; Pred. No. 2.9e-204;
Matches 488; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAAGACACACATAGAGCTTCTACTTCCATATCTCTAGCTTCTAAGAGAGCGTAT 60
DB 231 ATCAAGACACACATAGAGCTTCTACTTCCATATCTCTAGCTTCTAAGAGAGCGTAT 290
QY 61 TTAGCTGAAACACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAAC 120
DB 291 TTAGCTGAAACACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAAC 350
QY 121 CTTTGCAATTTGGCAGTCCTCTCTGGGGCGCTGAAGAGGCGGAGCCCAAGATGACCTT 180
DB 351 CTTTGCAATTTGGCAGTCCTCTCTGGGGCGCTGAAGAGGCGGAGCCCAAGATGACCTT 410
QY 181 CCAGATCTTTTCCAGCTTTCTGAGCCCAAGAAATAGACAAAGTGTCAATTTGGCAGA 240
DB 411 CCAGATCTTTTCCAGCTTTCTGAGCCCAAGAAATAGACAAAGTGTCAATTTGGCAGA 470
```

```
QY 241 CTGGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT 300
DB 471 CTGGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT 530
QY 301 TCTCCTGATCAGATGAACACTCACCTAAATTTAAGTTTTGAGCCTAACTTCTGCCAGAT 360
DB 531 TCTCCTGATCAGATGAACACTCACCTAAATTTAAGTTTTGAGCCTAACTTCTGCCAGAT 590
QY 361 AACCTCTGAAGTCCCACCAGCTCTAAAGAAAAGCCCCCAGAGGCAAAAAGGCCACAGTAT 420
DB 591 AACCTCTGAAGTCCCACCAGCTCTAAAGAAAAGCCCCCAGAGGCAAAAAGGCCACAGTAT 650
QY 421 TGTCTGAAAACCCAGTCCAAAAGATATTTTAAATTAAGGCTGCCGACTTCATTGGAAGAG 480
DB 651 TGTCTGAAAACCCAGTCCAAAAGATATTTTAAATTAAGGCTGCCGACTTCATTGGAAGAG 710
QY 481 CTATCCTCC 489
DB 711 CTATCCTCC 719
RESULT 6
ABA06731
ID ABA06731 standard; cDNA; 702 BP.
XX AC ABA06731;
XX DT 10-JAN-2002 (first entry)
XX DE Human cDNA SEQ ID NO: 397.
XX KW Human; gene therapy; neural disorder; immune system disorder;
XX KW muscular disorder; reproductive disorder; gastrointestinal disorder;
XX KW pulmonary disorder; cardiovascular disorder; renal disorder;
XX KW proliferative disorder; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200154474-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01349.
XX PR 31-JAN-2000; 2000US-179065P.
XX PR 04-FEB-2000; 2000US-180628P.
XX PR 24-FEB-2000; 2000US-184664P.
XX PR 02-MAR-2000; 2000US-186350P.
XX PR 16-MAR-2000; 2000US-189874P.
XX PR 17-MAR-2000; 2000US-190076P.
XX PR 18-APR-2000; 2000US-198123P.
XX PR 19-MAY-2000; 2000US-205515P.
XX PR 07-JUN-2000; 2000US-209467P.
XX PR 28-JUN-2000; 2000US-214886P.
XX PR 30-JUN-2000; 2000US-215135P.
XX PR 07-JUL-2000; 2000US-216647P.
XX PR 07-JUL-2000; 2000US-216880P.
XX PR 11-JUL-2000; 2000US-217487P.
XX PR 11-JUL-2000; 2000US-217496P.
XX PR 14-JUL-2000; 2000US-218290P.
XX PR 26-JUL-2000; 2000US-220963P.
XX PR 26-JUL-2000; 2000US-220964P.
XX PR 14-AUG-2000; 2000US-224518P.
XX PR 14-AUG-2000; 2000US-224519P.
XX PR 14-AUG-2000; 2000US-225213P.
XX PR 14-AUG-2000; 2000US-225214P.
XX PR 14-AUG-2000; 2000US-225266P.
XX PR 14-AUG-2000; 2000US-225267P.
XX PR 14-AUG-2000; 2000US-225268P.
XX PR 14-AUG-2000; 2000US-225270P.
XX PR 14-AUG-2000; 2000US-225270P.
XX PR 14-AUG-2000; 2000US-225475P.
XX PR 14-AUG-2000; 2000US-225477P.
```

PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 26-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.

PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR P-PSDB; ABB10509.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
PS Claim 1; SEQ ID NO: 397; 859pp + Sequence Listing; English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
SQ Sequence 702 BP; 179 A; 195 C; 187 G; 137 T; 4 other;

Query Match 8.8%; Score 348; DB 22; Length 702;
Best Local Similarity 99.5%; Pred. No. 4.4e-160;
Matches 638; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 3177 AGAAAGACAAAGCCCTACAGGAACGCTTTTCCGACCACTTTCCTGCAGGCTCC 3236
Db 2 AGAAAGACAAAGCCCTACAGGAACGCTTTTCCGACCACTTTCCTGCAGGCTCC 61
QY 3237 TGGGGATATGGTAGCTCATGAGGGGGCCCTCTGTCGGCTGGACTGTAAGTGAGTGGTTT 3296
Db 62 TGGGGATATGGTAGCTCATGAGGGGGCCCTCTGTCGGCTGGACTGTAAGTGAGTGGTTT 121
QY 3297 ACCGCCCCGGAGCTGACATGGCTACTCAATGCGCAACCTGTGCTACCAAGATGCCCTCCA 3356
Db 122 A-CGCCCCGGAGCTGACATGGCTACTCAATGCGCAACCTGTGCTACCAAGATGCCCTCCA 180

QY 3357 CAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCG 3416
Db 181 CAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCG 240
QY 3417 CGACGAGGACCTATAAGTGCATCGCTACCAACAACCGGCGAGAAATTCCTTTAGTCT 3476
Db 241 CGACG-AGGACCTATAAGTGCATCGCTACCAACAACCGGCGAGAAATTCCTTTAGTCT 299
QY 3477 GGAGCTCTCTGTAGTACCCAAAGAGTGAAGAACACCTGTGATCCTGGAGAACTACA 3536
Db 300 GGAGCTCTCTGTAGTACCCAAAGAGTGAAGAACACCTGTGATCCTGGAGAACTACA 359
QY 3537 GAATCGGGTGTCCCGAAGGCCACCCCGTGGAGCTGGAGTGGCGGTGATAGGCATGCC 3596
Db 360 GAATCGGGTGTCCCGAAGGCCACCCCGTGGAGCTGGAGTGGCGGTGATAGGCATGCC 419
QY 3597 CCCACCTGTCTTACTGGAAGAACACATGAGACCATCCCTTGCACACGAGAGGAT 3656
Db 420 CCCACCTGTCTTACTGGAAGAACACATGAGACCATCCCTTGCACACGAGAGGAT 479
QY 3657 CAGTATGCACCGACACACACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATC 3716
Db 480 CAGTATGCACCGACACACACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATC 539
QY 3717 AGACGGTGGATGTACACGTTGTGAGCCCAAGAAATGAAGCCGGCATCGTGTGCACTGC 3776
Db 540 AGACGGTGGATGTACACGTTGTGAGCCCAAGAAATGAAGCCGGCATCGTGTGCACTGC 599
QY 3777 CAGCGTGATATACGCTCAGTGGCACCATCAGATCCAC 3817
Db 600 CAGCGTGATATACGCTCAGTGGCACCATCAGATCCAC 640

RESULT 7
AAS28841
ID AAS28841 standard; cDNA; 702 BP.
AC AAS28841;
XX
XX
DT 07-NOV-2001 (first entry)
DE Human immunoglobulin encoding cDNA SEQ ID No 87.
XX
KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN WO200155315-A2.
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-0501326.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.

CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigene traits in a plant breeding program. (I) can be
 CC used to produce a tassel-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassel nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridisation or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridisation, and to identify the
 CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences.

XX
 SQ Sequence 281 BP; 54 A; 83 C; 91 G; 52 T; 1 other;

Query Match 0.5%; Score 21; DB 24; Length 281;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 GGAGCGGAGGCCAAGATGAC 177
 |||||
 Db 175 GGAGCGGAGGCCAAGATGAC 195

RESULT 12
 AAZ13687/C
 ID AAZ13687 standard; cDNA; 300 BP.
 AC AAZ13687;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:1156.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-494092/41.
 XX
 XX Novel human genes and their expression products which are
 XX differentially expressed in different cell types
 PT
 PS Claim 1; Page 924-925; 2479pp; English.

XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX
 SQ Sequence 300 BP; 99 A; 40 C; 63 G; 98 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2791 GAAATTCACATGATGAGAT 2810
 |||||
 Db 199 GAAATTCACATGATGAGAT 180

RESULT 13
 AAZ09013
 ID AAZ09013 standard; cDNA; 399 BP.
 AC AAZ09013;
 XX
 DT 19-OCT-1999 (first entry)
 XX
 DE Differentiation Induction Subtraction Hybridization (DISH)-75 sequence.
 XX
 KW DAP; differentiation-associated protein; terminal differentiation;
 KW growth arrest; differentiation induction subtraction hybridization;
 KW DISH; melanoma; breast; lung; colorectal; prostate; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9937774-A2.
 XX
 PD 29-JUL-1999.
 XX
 PF 25-JAN-1999; 99WO-US01549.
 XX
 PR 29-MAY-1998; 98US-0087167.
 PR 26-JAN-1998; 98US-0073298.
 PR 11-FEB-1998; 98US-0074441.
 PR 12-MAR-1998; 98US-0077804.
 PR 25-MAR-1998; 98US-0079326.
 PR 28-APR-1998; 98US-0083195.
 PR 15-MAY-1998; 98US-0085609.
 PR 26-MAY-1998; 98US-0086829.
 XX
 PA (GENQ-) GENQUEST INC.
 XX
 XX Fisher PB, Huang F;
 XX
 DR WPI; 1999-479051/40.
 XX
 XX Differentiation-associated proteins and related polynucleotides,
 XX useful for vaccine and pharmaceuticals to inhibit cell growth
 PT

XX Claim 1; Fig 8; 144pp; English.
 XX Sequences AA209006-209075 are Differentiation Induction Subtraction
 CC Hybridization (DISH) sequences, which encode Differentiation-Associated
 CC Proteins (DAPs). DAPs are associated with terminal differentiation and
 CC growth arrest and the sequences encoding them range from 97-903 base
 CC pairs in length. A DAP, a DAP fragment or a DAP polynucleotide may be
 CC useful in inhibiting the development of cancer including prostate,
 CC breast, lung and colorectal cancer, melanoma, astrocytoma or
 CC glioblastoma multiforme. Determining the level of a DAP or its coding
 CC sequence, in a tumour sample can be used to determine whether the tumour
 CC is malignant. The progression of cancer can be monitored by measuring DAP
 CC expression or activity levels over a period of time. An agent that
 CC increases expression of a DAP can also be used to inhibit the development
 CC of cancer.
 XX
 SQ Sequence 399 BP; 98 A; 91 C; 88 G; 122 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 399;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 CAAAAAAGTATTTTAAATA 457
 |||||
 DB 182 CAAAAAAGTATTTTAAATA 201

RESULT 14
 ABQ57885
 ID ABQ57885 standard; cDNA; 551 BP.
 XX
 AC ABQ57885;
 XX
 XX Homo sapiens.
 DT 02-AUG-2002 (first entry)
 XX
 DE Human colon cancer related nucleotide sequence SEQ ID NO:1580.
 KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200229086-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001WO-US30732.
 XX
 PR 02-OCT-2000; 2000US-237271P.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
 PI Thiagalingam A, Lewis ME;
 PI WPI; 2002-426115/45.
 XX
 PT New isolated nucleic acid that is differentially expressed in cancer
 PT tissues useful for determining the presence of colon cancer in a cell
 PT or tissue type, and in antisense therapy -
 XX
 PS Claim 1; Fig 1; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
 CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide
 CC encoded by (I) is useful for detecting cancer in a patient sample, and
 CC for detecting the presence or absence of a polynucleotide encoded by a
 CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived
 CC from (I) can be used for determining the presence of a nucleic acid which
 CC hybridises to (I), and for determining the phenotype of cells in a sample

CC of cells from a patient. (I) is useful for determining the presence of
 CC colon cancer in a cell or tissue type, for determining the presence or
 CC state of other type of cancer, in antisense therapy, to generate
 CC macroarrays on a solid surface, to identify a chromosome on which the
 CC corresponding gene resides, and in tissue profiling, forensics, genetic
 CC analysis, mapping and diagnostic applications. (I) can be used to raise
 CC antibodies, and to screen for peptide analogues and antagonists.
 XX

SQ Sequence 551 BP; 124 A; 126 C; 105 G; 194 T; 2 other;

Query Match 0.5%; Score 20; DB 24; Length 551;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 CAAAAAAGTATTTTAAATA 457
 |||||
 DB 324 CAAAAAAGTATTTTAAATA 343

RESULT 15
 AAZ15038/c
 ID AAZ15038 standard; cDNA; 733 BP.
 XX
 AC AAZ15038;
 XX
 XX Homo sapiens.
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:2507.
 XX

KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX
 XX Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX

PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX

XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 PT
 PS Claim 1; Page 1225; 2479pp; English.

XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and

CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists

Sequence 733 BP; 225 A; 142 C; 154 G; 205 T; 7 other;

Query Match	0.58;	Score 20;	DB 20;	Length 733;
Best Local Similarity	100.0%;	Pred. No. 44;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 2791 GAAATTCAACATGATGAGAT 2810
|||
Db 259 GAAATTCAACATGATGAGAT 240

Search completed: November 30, 2002, 14:15:10
Job time : 572 secs

c 977 15 0.4 1831 4 US-09-422-576D-1 Sequence 1, Appli
978 15 0.4 1839 1 US-08-442-248-3 Sequence 3, Appli
979 15 0.4 1839 1 US-08-440-815-3 Sequence 3, Appli
980 15 0.4 1839 4 US-08-486-449-3 Sequence 3, Appli
981 15 0.4 1841 2 US-08-820-521-1 Sequence 1, Appli
982 15 0.4 1841 4 US-09-248-715-1 Sequence 1, Appli
983 15 0.4 1841 4 US-09-248-715-1 Sequence 1, Appli
984 15 0.4 1842 4 US-08-961-527-85 Sequence 85, Appl
985 15 0.4 1843 1 US-08-774-169-2 Sequence 2, Appli
c 986 15 0.4 1843 1 US-08-499-523-9 Sequence 9, Appli
c 987 15 0.4 1843 3 US-09-128-345-9 Sequence 9, Appli
988 15 0.4 1854 5 PCT-US94-01101-1 Sequence 1, Appli
989 15 0.4 1859 4 US-09-286-529-19 Sequence 19, Appl
c 990 15 0.4 1859 4 US-09-286-529-19 Sequence 19, Appl
991 15 0.4 1869 4 US-09-406-045-2 Sequence 2, Appli
c 992 15 0.4 1869 1 US-09-406-045-2 Sequence 2, Appli
993 15 0.4 1880 1 US-08-247-475-49 Sequence 49, Appl
994 15 0.4 1880 1 US-08-479-650-49 Sequence 49, Appl
995 15 0.4 1880 1 US-08-191-866D-80 Sequence 80, Appl
996 15 0.4 1880 1 US-08-674-169-49 Sequence 49, Appl
997 15 0.4 1880 2 US-08-185-949B-80 Sequence 80, Appl
c 998 15 0.4 1944 4 US-09-149-476-144 Sequence 144, App
999 15 0.4 1962 4 US-08-791-115B-3 Sequence 3, Appli
c1000 15 0.4 1964 4 US-09-434-613-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-118-554-17
; Sequence 17, Application US/09118554A
; Patent No. 6365348
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS FOR DIAGNOSIS OF BREAST CANCER AND
; FILE REFERENCE: 210121.450C1
; CURRENT APPLICATION NUMBER: US/09/118,554A
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 08/988,255
; EARLIER FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (56)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (81)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (85)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (89)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (132)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (146)
; OTHER INFORMATION: Wherein n is a, c, g or t

; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (206)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (226)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (245)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (254)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (268)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (273)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (275)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (286)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (289)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (292)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (294)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (299)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (347)
; OTHER INFORMATION: Wherein n is a, c, g or t
US-09-118-554-17

Query Match 0.5%; Score 19; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAAGCCTCTAG 2293
Db 35 AGCAAGAAAGCCTCTAG 53
|||||

RESULT 2
US-09-118-627-17
; Sequence 17, Application US/09118627A
; Patent No. 6379951
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF BREAST CANCER
FILE REFERENCE: 210121.446C1
CURRENT APPLICATION NUMBER: US/09/118,627A
CURRENT FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 67
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 360
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(360)
OTHER INFORMATION: n = A,T,C or G
US-09-118-627-17

Query Match 0.5%; Score 19; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAGCCTCTTAG 2293
|||||
Db 35 AGCAAGAAGCCTCTTAG 53

RESULT 3
US-09-602-877A-17
Sequence 17, Application US/09602877A
Patent No. 6432707
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.446C5
CURRENT APPLICATION NUMBER: US/09/602,877A
CURRENT FILING DATE: 2000-06-22
NUMBER OF SEQ ID NOS: 107
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 360
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(360)
OTHER INFORMATION: n = A,T,C or G
US-09-602-877A-17

Query Match 0.5%; Score 19; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAGCCTCTTAG 2293
|||||
Db 35 AGCAAGAAGCCTCTTAG 53

RESULT 4
US-08-909-742-2/c
Sequence 2, Application US/08909742
Patent No. 6007991
GENERAL INFORMATION:
APPLICANT: Vimala S. Sivaraman
APPLICANT: Hsien-Yu Wang
APPLICANT: Craig C. Malbon
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES FOR MITOGEN-
TITLE OF INVENTION: ACTIVATED PROTEIN KINASES AS THERAPY FOR
TITLE OF INVENTION: BREAST CANCER
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann & Baron, LLP
STREET: 350 Jericho Turnpike
CITY: Jericho
STATE: New York
COUNTRY: USA
ZIP: 11753
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,742
FILING DATE: August 12, 1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/831,994
FILING DATE: April 1, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,520
FILING DATE: March 28, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Adams, Lindsay S.
REGISTRATION NUMBER: 36,425
REFERENCE/DOCKET NUMBER: 178-225 CIP II
TELEPHONE: (516) 822-3550
TELEFAX: (516) 822-3582
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1611 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-909-742-2

Query Match 0.5%; Score 19; DB 3; Length 1611;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1499 TGGTCATTGCTGAGGTGTT 1517
|||||
Db 701 TGGTCATTGCTGAGGTGTT 683

RESULT 5
US-09-412-289-2/c
Sequence 2, Application US/09412289
Patent No. 6271210
GENERAL INFORMATION:
APPLICANT: Sivaraman, Vimala S.
APPLICANT: Wang, Hsien-yu
APPLICANT: Malbon, Craig C.
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES FOR MITOGEN-ACTIVATED
TITLE OF INVENTION: PROTEIN KINASES AS THERAPY FOR CANCER
FILE REFERENCE: Seq. 1-4 (178-225 CIP II/CON)
CURRENT APPLICATION NUMBER: US/09/412,289
CURRENT FILING DATE: 1999-10-05
EARLIER APPLICATION NUMBER: 08/909,742
EARLIER FILING DATE: 1997-08-12
EARLIER APPLICATION NUMBER: 08/831,994
EARLIER FILING DATE: 1997-04-01
EARLIER APPLICATION NUMBER: 08/827,520
EARLIER FILING DATE: 1997-03-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1611
TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: blank
US-09-412-289-2

Query Match 0.5%; Score 19; DB 4; Length 1611;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1499 TGGTCATTGCTGAGGCTTT 1517
Db 701 TGGTCATTGCTGAGGCTTT 683

RESULT 6
US-07-938-333A-3
; Sequence 3, Application US/07938333A
; Patent No. 5439822
; GENERAL INFORMATION:
; APPLICANT: Katsumata, Ryoichi
; APPLICANT: Takano, Yutaka
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.50inch
; COMPUTER: Toshiba J-3100 GT (IBM PC compatible)
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Apple Macintosh Hypercard 1.2.2,
; SOFTWARE: Apple File Exchange J1-1.1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/938,333A
; FILING DATE: 19920828
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lippert, Neils T.
; REGISTRATION NUMBER: 25,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-819-8200
; TELEFAX: 212-354-8113
; TELEX: 126201
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2135 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13032
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 514 to 1806
; IDENTIFICATION METHOD: E

US-07-938-333A-3
Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGTCTG 854
Db 1504 AAGTCCAGAAGGAAGTCTG 1522

RESULT 7
US-07-938-333A-3
; Sequence 3, Application US/07938333A
; Patent No. 5439822
; GENERAL INFORMATION:
; APPLICANT: Katsumata, Ryoichi
; APPLICANT: Takano, Yutaka
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.50inch
; COMPUTER: Toshiba J-3100 GT (IBM PC compatible)
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Apple Macintosh Hypercard 1.2.2,
; SOFTWARE: Apple File Exchange J1-1.1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/938,333A
; FILING DATE: 19920828
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lippert, Neils T.
; REGISTRATION NUMBER: 25,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-819-8200
; TELEFAX: 212-354-8113
; TELEX: 126201
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2135 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13032
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 514 to 1806
; IDENTIFICATION METHOD: E

US-07-938-333A-3
Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGTCTG 854
Db 1504 AAGTCCAGAAGGAAGTCTG 1522

RESULT 7

US-08-660-216A-3
; Sequence 3, Application US/08660216A
; Patent No. 5700661
; GENERAL INFORMATION:
; APPLICANT: Katsumata, Ryoichi
; APPLICANT: Takano, Yutaka
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.50inch
; COMPUTER: Toshiba J-3100 GT (IBM PC compatible)
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Apple Macintosh Hypercard 1.2.2,
; SOFTWARE: Apple File Exchange J1-1.1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,216A
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/398,456
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,333
; FILING DATE: 28-AUG-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lippert, Neils T.
; REGISTRATION NUMBER: 25,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-819-8200
; TELEFAX: 212-354-8113
; TELEX: 126201
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2135 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13032
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 514 to 1806
; IDENTIFICATION METHOD: E

US-08-660-216A-3
Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGTCTG 854
Db 1504 AAGTCCAGAAGGAAGTCTG 1522

RESULT 8
US-08-211-430-1
; Sequence 1, Application US/08211430
; Patent No. 5763166
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC SEQUENCE OF THE GENE ASSOCIATED WITH
; TITLE OF INVENTION: X CHROMOSOME LINKED KALLMANN SYNDROME, CORRESPONDING

US-08-211-430-1
Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGTCTG 854
Db 1504 AAGTCCAGAAGGAAGTCTG 1522

RESULT 8
US-08-211-430-1
; Sequence 1, Application US/08211430
; Patent No. 5763166
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC SEQUENCE OF THE GENE ASSOCIATED WITH
; TITLE OF INVENTION: X CHROMOSOME LINKED KALLMANN SYNDROME, CORRESPONDING

US-08-211-430-1
Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGTCTG 854
Db 1504 AAGTCCAGAAGGAAGTCTG 1522

;; TITLE OF INVENTION: PEPTIDE SEQUENCES, DIAGNOSTIC APPLICATIONS.

;; NUMBER OF SEQUENCES: 32

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/211,430

;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 6314 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; MOLECULE TYPE: cDNA

;; HYPOTHETICAL: NO

;; ANTI-SENSE: NO

;; ORIGINAL SOURCE:

;; ORGANISM: Homo sapiens

;; CELL LINE: foetal brain cells

;; US-08-211-430-1

Query Match 0.5%; Score 19; DB 1; Length 6314;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3515 CTGTGATCCTGGAGAAACT 3533

Db 1223 CTGTGATCCTGGAGAAACT 1241

RESULT 9

US-07-988-273-1

;; Sequence 1, Application US/07988273

;; Patent No. 5512434

;; GENERAL INFORMATION:

;; APPLICANT: AARONSON, Stuart A.

;; APPLICANT: BOTTARO, Donald P.

;; APPLICANT: ISHIBASHI, Toshio

;; APPLICANT: MIKI, Toru

;; TITLE OF INVENTION: EXPRESSION CLONING OF A HUMAN

;; TITLE OF INVENTION: PHOSPHATASE

;; NUMBER OF SEQUENCES: 7

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Foley & Lardner

;; STREET: 3000 K Street, N.W., Suite 500

;; CITY: Washington, D.C.

;; COUNTRY: USA

;; ZIP: 20007-5109

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/07/988,273

;; FILING DATE: 19921214

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: BENT, Stephen A.

;; REGISTRATION NUMBER: 29,768

;; REFERENCE/DOCKET NUMBER: 40399/182 NIHD

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (202)672-5300

;; TELEFAX: (202)672-5399

;; TELEX: 904136

;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 861 base pairs

;; TYPE: NUCLEIC ACID

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 29..586

;; US-07-988-273-1

Query Match

Best Local Similarity 0.5%; Score 18; DB 1; Length 861;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 GGCTGCCGACTTCATTGA 476

Db 340 GGCTGCCGACTTCATTGA 357

RESULT 10

PCT-US93-12019-1

;; Sequence 1, Application PC/TUS9312019

;; GENERAL INFORMATION:

;; APPLICANT:

;; TITLE OF INVENTION: EXPRESSION CLONING OF A HUMAN

;; TITLE OF INVENTION: PHOSPHATASE

;; NUMBER OF SEQUENCES: 7

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: PCT/US93/12019

;; FILING DATE:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/988,273

;; FILING DATE: 14-DEC-1992

;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 861 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 29..586

;; PCT-US93-12019-1

Query Match

Best Local Similarity 0.5%; Score 18; DB 5; Length 861;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 GGCTGCCGACTTCATTGA 476

Db 340 GGCTGCCGACTTCATTGA 357

RESULT 11

US-09-054-680-3

;; Sequence 3, Application US/09054680

;; Patent No. 6387694

;; GENERAL INFORMATION:

;; APPLICANT: McKinney, John D.

;; APPLICANT: Jacobs Jr., William R.

;; TITLE OF INVENTION: MYCOBACTERIAL ISOCITRATE LYASE GENE AND USES THEREOF

;; FILE REFERENCE: 96700/474

;; CURRENT APPLICATION NUMBER: US/09/054,680

;; CURRENT FILING DATE: 1998-04-03

;; NUMBER OF SEQ ID NOS: 5

;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 3

;; LENGTH: 1287

;; TYPE: DNA

;; ORGANISM: Mycobacterium smegmatis

;; FEATURE:

;; NAME/KEY: gene

;; LOCATION: (1)..(1287)

US-09-054-680-3

Query Match 0.5%; Score 18; DB 4; Length 1287;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 835 GAAGTTCAGAAGAACT 852
|||||
Db 990 GAAGTTCAGAAGAACT 1007

RESULT 12

US-08-743-637B-20
; Sequence 20, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
US-08-743-637B-20

Query Match 0.5%; Score 18; DB 2; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2396 TTTCATCCCGAGGAA 2413
|||||
Db 104 TTTCATCCCGAGGAA 121

RESULT 13

US-08-526-840B-20

; Sequence 20, Application US/08526840B
; Patent No. 6001564

; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
US-08-526-840B-20

Query Match 0.5%; Score 18; DB 3; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2396 TTTCATCCCGAGGAA 2413
|||||
Db 104 TTTCATCCCGAGGAA 121

RESULT 14
US-08-748-170A-3
; Sequence 3, Application US/08748170A
; Patent No. 6048688
; GENERAL INFORMATION:
; APPLICANT: Korth, Kevin G.
; APPLICANT: Heathcock, Sarah E.
; APPLICANT: Huard, Linda S.
; TITLE OF INVENTION: Method for Detecting Pseudomonas
; TITLE OF INVENTION: aeruginosa Using Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia

; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,170A
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,697
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 11302-0350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..2760
; OTHER INFORMATION: /note= "Found within the published
; OTHER INFORMATION: sequence of the P. aeruginosa Exotoxin A gene; GenBank access
; OTHER INFORMATION: number K01397; NCBI sequence ID: 151215"
US-08-748-170A-3

Query Match 0.5%; Score 18; DB 3; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2396 TTTCATCCCGCGGAA 2413
Db 104 TTTCATCCCGCGGAA 121

RESULT 15

US-09-047-148-1
; Sequence 1, Application US/09047148
; Patent No. 6086900
; GENERAL INFORMATION:
; APPLICANT: Draper, Rockford
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR USING
; TITLE OF INVENTION: MEMBRANE-PENETRATING PROTEINS TO CARRY MATERIALS ACROSS
; TITLE OF INVENTION: CELL MEMBRANES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/047,148

; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,056
; FILING DATE: 26-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSP:072
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-047-148-1

Query Match 0.5%; Score 18; DB 3; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2396 TTTCATCCCGCGGAA 2413
Db 104 TTTCATCCCGCGGAA 121

Search completed: November 30, 2002, 17:05:28
Job time : 162 secs

c 969	15	0.4	458	10	US-09-864-761-1187	Sequence 1187, Ap
c 970	15	0.4	458	10	US-09-867-701-729	Sequence 729, App
c 971	15	0.4	459	10	US-09-864-761-3743	Sequence 3743, Ap
c 972	15	0.4	460	10	US-09-864-761-1245	Sequence 1245, Ap
c 973	15	0.4	460	10	US-09-954-456-1096	Sequence 1096, Ap
c 974	15	0.4	462	10	US-09-287-070-3	Sequence 3, Appl
c 975	15	0.4	463	10	US-09-864-761-6404	Sequence 6404, Ap
c 976	15	0.4	463	10	US-09-864-761-16294	Sequence 16294, A
c 977	15	0.4	464	10	US-09-770-444-353	Sequence 353, App
c 978	15	0.4	464	10	US-09-864-761-14124	Sequence 14124, A
c 979	15	0.4	464	10	US-09-960-352-5924	Sequence 5924, Ap
c 980	15	0.4	465	10	US-09-864-761-1396	Sequence 1396, Ap
c 981	15	0.4	465	10	US-09-864-761-4006	Sequence 4006, Ap
c 982	15	0.4	465	10	US-09-864-761-11249	Sequence 11249, A
c 983	15	0.4	465	10	US-09-864-761-17469	Sequence 17469, A
c 984	15	0.4	466	10	US-09-864-761-5761	Sequence 5761, Ap
c 985	15	0.4	466	10	US-09-783-590-10805	Sequence 10805, A
c 986	15	0.4	467	10	US-09-770-444-274	Sequence 274, App
c 987	15	0.4	467	10	US-09-864-761-16108	Sequence 16108, A
c 988	15	0.4	468	10	US-09-770-444-250	Sequence 250, App
c 989	15	0.4	468	10	US-09-833-381-257	Sequence 257, App
c 990	15	0.4	468	10	US-09-924-035A-210	Sequence 210, App
c 991	15	0.4	470	10	US-09-864-761-5836	Sequence 5836, Ap
c 992	15	0.4	470	10	US-09-864-761-14044	Sequence 14044, A
c 993	15	0.4	470	10	US-09-867-701-1599	Sequence 1599, Ap
c 994	15	0.4	471	10	US-09-815-242-2901	Sequence 2901, Ap
c 995	15	0.4	471	10	US-09-917-800A-910	Sequence 910, App
c 996	15	0.4	471	10	US-09-833-381-1387	Sequence 1387, Ap
c 997	15	0.4	471	10	US-09-867-701-791	Sequence 791, App
c 998	15	0.4	472	10	US-09-770-444-146	Sequence 146, App
c 999	15	0.4	472	10	US-09-864-761-3901	Sequence 3901, Ap
c1000	15	0.4	472	10	US-09-864-761-4564	Sequence 4564, Ap

ALIGNMENTS

RESULT 1
US-09-764-853-218
; Sequence 218, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 2059
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-218

Query Match	19.9%	Score 787;	DB 10;	Length 2059;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 787;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 3177	AGAAAGACAAAGAGCCCTACAGGACGCTTTTCCGACACCATTTCTCGCAGGCTCC	3236		
Db 9	AGAAAGACAAAGAGCCCTACAGGACGCTTTTCCGACACCATTTCTCGCAGGCTCC	68		
Qy 3237	TGGGGATATGCTAGCTCATGAGGGCCCTCTGTCGCTGGACTGTAAGTGAAGTGT	3296		
Db 69	TGGGGATATGCTAGCTCATGAGGGCCCTCTGTCGCTGGACTGTAAGTGAAGTGT	128		
Qy 3297	ACCCGCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCCA	3356		
Db 129	ACCCGCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCCA	188		
Qy 3357	CAAGATGCTGTCAGGGAGACCGGAGTCCACTCTCTGCTCAATTGACCCACTCACTACGC	3416		

Db 189	CAAGATGCTGCTGAGGAGACCGAGTCCACTCTCTCTCTCATTGACCCACTCACTACGC	248
Qy 3417	CGAGCGAGGACCTATAAGTGCATCGTACCAACAAACCGGCGAGAAATCTTTTAGTCT	3476
Db 249	CGAGCGAGGACCTATAAGTGCATCGTACCAACAAACCGGCGAGAAATCTTTTAGTCT	308
Qy 3477	GGAGCTCTCTGTAGTACGCCAAAGAGGTGAAGAAAGCACTGTGTATCTCGAGAACTACA	3536
Db 309	GGAGCTCTCTGTAGTACGCCAAAGAGGTGAAGAAAGCACTGTGTATCTCGAGAACTACA	368
Qy 3537	GAACTGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCC	3596
Db 369	GAACTGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCC	428
Qy 3597	CCACCTGCTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGCACAGAGAGGAT	3656
Db 429	CCACCTGCTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGCACAGAGAGGAT	488
Qy 3657	CAGTATCACCAGGACACAAACAGGGTATGCTGCTTCTCTCATTGACGCCAGCAAGAAATC	3716
Db 489	CAGTATCACCAGGACACAAACAGGGTATGCTGCTTCTCTCATTGACGCCAGCAAGAAATC	548
Qy 3717	AGAGCGTGGATGTTACAGTTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGACTGC	3776
Db 549	AGAGCGTGGATGTTACAGTTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGACTGC	608
Qy 3777	CAGGCTGATATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGCTGTCCGGCC	3836
Db 609	CAGGCTGATATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGCTGTCCGGCC	668
Qy 3837	CAGTGGCAGTCCGCTACGGATCTCTCACCAGTAAAGGACTTTGACATATTTTCTGCTTTTC	3896
Db 669	CAGTGGCAGTCCGCTACGGATCTCTCACCAGTAAAGGACTTTGACATATTTTCTGCTTTTC	728
Qy 3897	CTCCATGGAAGCAGCATGCTGTATTCATGCTCTTCTCGGAGTGTAGTGAGAGTGATGA	3956
Db 729	CTCCATGGAAGCAGCATGCTGTATTCATGCTCTTCTCGGAGTGTAGTGAGAGTGATGA	788
Qy 3957	ACTTTAA 3963	
Db 789	ACTTTAA 795	

RESULT 2
US-09-764-853-397
; Sequence 397, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 397
; LENGTH: 702
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (587)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: SITE
; LOCATION: (661)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: SITE
; LOCATION: (670)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-764-853-397

```
Query Match      8.8%; Score 348; DB 10; Length 702;
Best Local Similarity 99.5%; Pred. No. 4.1e-172;
Matches 638; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 3177 AGAAGAGACAAAGAGCCCTACAGGAAGCGCTTTTCCGACCACATTTCTCGAGGCTCC 3236
Db      |||
QY 3237 TGGGATATGGTAGCTCATGAGGGCGCCTCTCTCGGCTGGACTGTAAGGTGAGTGTTT 3296
Db      |||
QY 3297 ACCGCCCGGAGCTGACATGGCTACTCAATGCGCCAACTGTCTACAGATGCCCTCCCA 3356
Db      |||
QY 3357 CAAGATGCTGGTCAGGAGACCGGAGTCCACITCTCTGCTCATTTGACCCACTCACTCAGCG 3416
Db      |||
QY 3417 CGACGAGGACCTATAAGTGCATCGCTACCAACAAACCGGGCAGAAATCTTTTAGTCT 3476
Db      |||
QY 3477 GGAGCTCTCTGTAGTAGCAAGAGAGTGAAGAAAGCACCTGTGATCCTGGAGAAACTACA 3536
Db      |||
QY 3537 GAACCTGGTGTTCGGAAGCCACCCCTGGAGACTGGAGTGGCGGTGATGGCATGCC 3596
Db      |||
QY 3597 CCACCTGTGTCTACTGGAAGAAAGCAATGAGACCATCCCTTGCACCAAGAGAGGAT 3656
Db      |||
QY 420 CCCACCTGTGTCTACTGGAGAAAGCAATGAGACCATCCCTTGCACCAAGAGAGGAT 479
QY 3657 CAGTATGCACAGGACACACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATC 3716
Db      |||
QY 480 CAGTATGCACAGGACACACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATC 539
QY 3717 AGACGCTGGATGTACACCTGTGCAGCCAGAAATGAAGCGGCGATGCTGCTGCTACTGC 3776
Db      |||
QY 540 AGACGCTGGATGTACACCTGTGCAGCCAGAAATGAAGCGGCGATGCTGCTGCTACTGC 599
QY 3777 CAGGCTGGATATATACGCTCAGTGGCACCACCATCAGATCCCAAC 3817
Db      |||
QY 600 CAGGCTGGATATATACGCTCAGTGGCACCACCATCAGATCCCAAC 640

RESULT 3
US-09-563-817-656
; Sequence 656, Application US/09563817
; Patent No. US20020095031A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020095031A1el Human Polynucleotides and the
; FILE REFERENCE: LEX-0021-USA
; CURRENT APPLICATION NUMBER: US/09/563,817
; PRIOR FILING DATE: 2001-06-11
; PRIOR FILING DATE: 1999-05-04
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 656
; LENGTH: 355
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(355)
```

```
; OTHER INFORMATION: n = A,T,C or G
US-09-563-817-656

Query Match      4.5%; Score 177; DB 10; Length 355;
Best Local Similarity 99.6%; Pred. No. 1.9e-82;
Matches 227; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 903 GTGGTACTGTGAAGGCAAGGAGCTTGAATAATCCCGAGATATTCACATCTCCAGGCAGG 962
Db      |||
QY 963 AAATCTGCATCTACTGACCAATTCGGAAGCCTTTGAAGAGGACACAGGACGCTATTCTCTG 1022
Db      |||
QY 1023 CTTTGGCTTCTAAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAGGGGT 1082
Db      |||
QY 121 CTTTGGCTTCTTACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAGGGGT 180
QY 1083 TTTCTTCTTCTGACTCAGAAGCGGACCCCTAACAGGAAGAGATGAATCG 1130
Db      |||
QY 181 TTTCTTCTTCTGACTCAGAAGCGGACCCCTAACAGGAAGAGATGAATCG 228

RESULT 4
US-09-294-093B-4682
; Sequence 4682, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalquidi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4682
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700354793H1
; LOCATION: 127
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-4682

Query Match      0.5%; Score 21; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 GGAGGGCGGAGGCCAAGATGAC 177
Db      |||
QY 175 GGAGGGCGGAGGCCAAGATGAC 195

RESULT 5
US-09-783-590-7307/C
; Sequence 7307, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
```

; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7307
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (58)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (107)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (152)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (175)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-7307

Query Match 0.5%; Score 19; DB 10; Length 202;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2225 CTCGGGTGTCACCTTGAG 2243
|||||
Db 129 CTCGGGTGTCACCTTGAG 111

RESULT 6
US-09-745-288-17
; Sequence 17, Application US/09745288
; Patent No. US20010018058A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.44601
; CURRENT APPLICATION NUMBER: US/09/745,288
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(360)
; OTHER INFORMATION: n = A,T,C or G
US-09-745-288-17

Query Match 0.5%; Score 19; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAAGCCTCTTAG 2293
|||||
Db 35 AGCAAGAAAGCCTCTTAG 53

RESULT 7
US-09-974-300-244
; Sequence 244, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-244

Query Match 0.5%; Score 19; DB 10; Length 1228;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 984 TCGGGAAGCCTTTGAAGAG 1002
|||||
Db 219 TCGGGAAGCCTTTGAAGAG 237

RESULT 8
US-09-974-298-31
; Sequence 31, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1382961.3
US-09-974-298-31

Query Match 0.5%; Score 19; DB 9; Length 1645;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1627 TCTCTTCACTCAGCCTACT 1645
|||||
Db 25 TCTCTTCACTCAGCCTACT 43

RESULT 9
US-09-919-172-1
; Sequence 1, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS

FILE REFERENCE: PA-0036 US
CURRENT APPLICATION NUMBER: US/09/919,172
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/222,469
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 1645
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020119463A1 1382961.3
US-09-919-172-1

Query Match 0.5%; Score 19; DB 10; Length 1645;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1627 TCTCTTCACTCAGCCAACT 1645
|||||
DB 25 TCTCTTCACTCAGCCAACT 43

RESULT 10
US-09-925-299-231
Sequence 231, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 231
LENGTH: 1823
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1593)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1714)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-231

Query Match 0.5%; Score 19; DB 10; Length 1823;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3737 TGTGAGCCCAAGAAATGAAGC 3755
|||||
DB 1408 TGTGAGCCCAAGAAATGAAGC 1426

RESULT 11
US-09-799-983-3
Sequence 3, Application US/09799983
Patent No. US20010029033A1
GENERAL INFORMATION:
APPLICANT: Shami, Paul
APPLICANT: Parker, Charles

TITLE OF INVENTION: NOVEL GENE RNO UPREGULATED IN LEUKEMIA CELLS BY NITRIC OXIDE
FILE REFERENCE: 1321.2.51
CURRENT APPLICATION NUMBER: US/09/799,983
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: US 60/186,971
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 2494
TYPE: DNA
ORGANISM: Homo sapiens
US-09-799-983-3

Query Match 0.5%; Score 19; DB 10; Length 2494;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2797 CAACATGATGAGATCCCA 2815
|||||
DB 442 CAACATGATGAGATCCCA 460

RESULT 12
US-09-799-983-5
Sequence 5, Application US/09799983
Patent No. US20010029033A1
GENERAL INFORMATION:
APPLICANT: Shami, Paul
APPLICANT: Parker, Charles
TITLE OF INVENTION: NOVEL GENE RNO UPREGULATED IN LEUKEMIA CELLS BY NITRIC OXIDE
FILE REFERENCE: 1321.2.51
CURRENT APPLICATION NUMBER: US/09/799,983
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: US 60/186,971
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 2511
TYPE: DNA
ORGANISM: Homo sapiens
US-09-799-983-5

Query Match 0.5%; Score 19; DB 10; Length 2511;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2797 CAACATGATGAGATCCCA 2815
|||||
DB 459 CAACATGATGAGATCCCA 477

RESULT 13
US-09-799-983-1
Sequence 1, Application US/09799983
Patent No. US20010029033A1
GENERAL INFORMATION:
APPLICANT: Shami, Paul
APPLICANT: Parker, Charles
TITLE OF INVENTION: NOVEL GENE RNO UPREGULATED IN LEUKEMIA CELLS BY NITRIC OXIDE
FILE REFERENCE: 1321.2.51
CURRENT APPLICATION NUMBER: US/09/799,983
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: US 60/186,971
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2682
TYPE: DNA
ORGANISM: Homo sapiens
US-09-799-983-1

```
Query Match      0.5%; Score 19; DB 10; Length 2682;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2797 CAACATGATGAGATCCCA 2815
   |||||
Db 459 CAACATGATGAGATCCCA 477

RESULT 14
US-09-815-242-7797
; Sequence 7797, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7797
; LENGTH: 3447
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(3447)
US-09-815-242-7797

Query Match      0.5%; Score 19; DB 10; Length 3447;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 CTGGGGCAACCTCCCGGT 811
   |||||
Db 51 CTGGGGCAACCTCCCGGT 69

RESULT 15
US-09-954-456-319
; Sequence 319, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 319
; LENGTH: 6314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-319

Query Match      0.5%; Score 19; DB 10; Length 6314;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3515 CTGTGATCCTGGAGAAACT 3533
   |||||
Db 1223 CTGTGATCCTGGAGAAACT 1241

Search completed: November 30, 2002, 18:41:22
Job time : 185 secs
```


AF328296 AF328296 5707 bp mRNA linear PRI 06-MAY-2001
LOCUS Homo sapiens myopalladin mRNA, complete cds.
DEFINITION AF328296
ACCESSION AF328296
VERSION AF328296.1 GI:13957726
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5707)
AUTHORS Bang,M.L., Mudry,R.E., McElhinny,A.S., Trombitas,K., Geach,A.J.,
Yamasaki,R., Sorimachi,H., Granzler,H., Gregorio,C.C. and Labelt,S.
TITLE Myopalladin, a novel 145-kilodalton sarcomeric protein with
multiple roles in z-disc and I-band protein assemblies
J. Cell Biol. 153 (2), 413-427 (2001)
JOURNAL 21206024
MEDLINE
PUBMED 11309420
REFERENCE 2 (bases 1 to 5707)
AUTHORS Bang,M.-L.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2000) Structural and Computational Biology,
European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1,
Heidelberg 69117, Germany
FEATURES
source
1..5707
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q21.1"
/tissue_type="striated muscle"
488..4450
/codon_start=1
/product="myopalladin"
/protein_id="AAK50625.1"
/db_xref="GI:13957727"
/translation="MODSIEASTSISQLRESYLAEATHRRGNERSRAPSSNPCH
GSPGAAAGGGGDDLPDLFAFLSEELDESVALRLAIINYPLEKADETQARKRLSP
DOMKHNPLSGFCDONPRSTSSKESQEAARPOYCYCSETQSKVFLNKAADFTIE
LSLFLKSHKRIIPRACKNHKSKLESONKMOENSSFDLSERRRSPVPIPAD
TRONEVNHLEQOEAKREAEAAAGDTPGSSPSLSLYEEELGQPPRTOKLR
SREVPETRVQLDCIVGIPPOVRWYCEGKELENSDPDIHVQAGNLHLTIAEPEE
DTGRYSKALVGTDSAEIYIEGVSSESDGPNKEEMRIQKPNVSSPPTTSA
VIVPAVQAQHLVADPRVATIQCCSPTNYLQGLDGKPIIAAPVFTKMLQNLASSEQ
LVVFECKVAGSPKVVYREGTLIEDSPDFRILOKPRSMAPETICTLVIAEVFAE
DSGCTCTAGNKYGTSSIAQLHVRGNEDLSNGLSHANSTTNLAAIEPQPPHSE
PPSVEQPPKPLEGLVNHNEPRSSRIGLRVHNLPEDDKGSSEAEAGVTTROT
PDSEFQENQAKTPEPSPVKEPPVLAAPKLDLSTQLQLHNVLLERQHLQNP
SSPEFPFNMTVNSNAPPVATTSNKQVKAPSSQTFSLARPKYFFFTNTAATVAPS
SSPVFTLSSTVQIQTYSKESLLVSHPSVQTKSPGSLSIONEPLPPGTEPTPPPT
FSIPRGNOFCVSPVPSPTSRIQNPVAFLSVLPSLPAIPTNAMGPRSPMP
SIOGLAKNTKSPQVNDNTRETKNAVIRDLGKITEFSDVRPNQOEYKISSEFQRLWN
EIEFLERTPVDESDDIOHDEIPTGKCIAPFDKRLKHFRTVEGSPVTTCKIVGP
VPKYWEKDKQISKRENECHKMRREGDGTCSLHESITSDDDGNITYMAANPOGRISC
SGHLMVSLPIRLTYSAGHSRQVRDEKPELQREFRPHFLQADQMDVAHEGR
LCRLDKVSLPPELTPMLLNGQVPLVDASHKMLVRETGVHSLDPLTDQMDAGTYK
IANKTQNSFELSLSVAAKEVKAPVLEILQKQNGVPEGHVRLCRVTGMPPPVFY
WKQNETIPTCRISMHODITGVACLLIQAPAKKSDAGVWTLAKNAGIVSTCARD
IYAQWHHQIIPPMVSRPSCSGSLTSKGLDIFSAFSSMESTWYSCSSRVVESDEL
1277..1531
/note="Region: Ig-I immunoglobulin domain"
misc_feature
1790..2044
/note="Region: Ig-I immunoglobulin domain"
misc_feature
3299..3565
/note="Region: Ig-I immunoglobulin domain"
misc_feature
3722..3946
/note="Region: Ig-I immunoglobulin domain"
misc_feature
4001..4246
/note="Region: Ig-I immunoglobulin domain"
BASE COUNT 1623 a 1437 c 1294 g 1353 t
ORIGIN

Alignment Scores:
Pred. No.: 2.06e-252 Length: 5707
Score: 6869.00 Matches: 1314
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 6
Query Match: 99.70% Indels: 0
DB: 9 Gaps: 0
US-09-818-990B-2 (1-1320) x AF328296 (1-5707)
Qy 1 MetGlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 488 ATGCAAGACGACATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 547
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 548 TTAGCTGNAACACACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAC 607
Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlyGlnAspLeu 60
Db 608 CTTTGCCATTTCGCGAGCTCTTCTGGGCGCGCTGAAGGAGCGGAGGCAAGATGACCTT 667
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 668 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAATTAGACGAAAGTGTCAATTTGGCA 727
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 728 CTGGCCATCAATTAGACCCCTTTTGAGAGAGCGACATGAACTCAAGCTAGAAAACGACTT 787
Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 788 TCTCCTGATCAGATGAACACACTCACCTAAATTAAGTTTGTAGCCTTAACCTTCCAGGAT 847
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Db 848 AACCTCTGAAGTCCACACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT 907
Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 908 TGTTCGAAACCCAGTCCAAAAAGTATTTTAATAAGGCTGCCGACTTCATTGAAGAG 967
Qy 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
Db 968 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGCTGCCTCAAAAA 1027
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSerPheSer 200
Db 1028 CACAAGAGTAAACTGGGAATCTCAAAACAAAGTTTATGCAGGAAAAACAGCTCCAGTTTCTCA 1087
Qy 201 AspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThrArg 220
Db 1088 GATCTGTGAGAAACAGCAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCA 1147
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
Db 1148 GATAATGAAGTGAATCACGCCCTCGAACAGCAGGAAAGCAAGAGCGGTGAAGCGGAGCAG 1207
Qy 241 AlaAlaSerGluAlaAlaGluGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Db 1208 GCTGCCAGTAGGCGGCTGGTGGAGACACTACACACAGGGTCTTCCCTTCATCTCTGTAC 1267
Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1268 TATGAGAACCTCTGGGGCAACCTCCCGGTTCACTCAAAAGTTACGGAGCAGAGAGATT 1327
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
Db 1328 CCAGAAGGAACCTCGAGTACAGTTGGATTGGCATAGTGGTAGGAATTCACCAACCTCAAGTA 1387
Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
Db 1388 AGGTGGTACTGTGAAGCGAAGGAGCTTTGAAAATTTCCCCAGATATTCACATCTGTCAGGCA 1447


```
|||||
Db 3608 ATTCCAGTCGCTAACTCTGCTGGTCAGTCTCACAGGGAAGATCCGAGTCAAGAA 3667
Qy 1061 ArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGly 1080
Db 3668 AGACAAAGAGCCCCACAGAAACGCTTTTCGACACACATTTCCGACGCTCCCTGGG 3727
Qy 1081 AspMetValAlaHisGluGluArgLeuGluCysArgLeuAspCysLysValSerGlyLeuPro 1100
Db 3728 GATATGTAGTCTCATGAGGGCGCTCTGTCGGCTGACATGTAGGTGAGTGGTTTACCG 3787
Qy 1101 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLys 1120
Db 3788 CCCCAGGAGTGCATGCTACTCAATGGCCAACTGTGCTACCCAGATGCTCCGCCCAAG 3847
Qy 1121 MetLeuValArgGluThrGlyValHisSerLeuLeuLeuLeuLeuLeuLeuLeuLeu 1140
Db 3848 ATGCTGTCAGGAGAGCGGAGTGCCTCTCTGCTCATTTGACCCCACTCACTCAGCGCGAC 3907
Qy 1141 AlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGlu 1160
Db 3908 GCAGGACCTATAGTGCATGCTCCTACCAAAACCGGCGAGAAATCTTTTAGTCTGGAG 3967
Qy 1161 LeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsn 1180
Db 3968 CTCCTCTAGTACCAAGAGGTCAAGAAAGCACTCTGATCCTGGAGAACTACAGAAC 4027
Qy 1181 CysGlyValProGluGluHisProValArgLeuGluCysArgValIleGlyMetProPro 1200
Db 4028 TGGCGGTCTCCGAGAGCCACCCGCTGAGAGTGGAGTGGCGGTGATGAGCATGCCCCA 4087
Qy 1201 ProValPheTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSer 1220
Db 4088 CCTGTCTTACTCGAGAGAAGACAATGAGACCATCCTTGCACACAGAGAGATCAGT 4147
Qy 1221 MethisGlnAspThrThrGlyTyrAlaCysLeuLeuLeuGlnProAlaLysLysSerAsp 1240
Db 4148 ATGCACAGGACACACAGGATGCTGCTGCTTCTCATTCAGCCAGCCCAAGAAATCAGAC 4207
Qy 1241 AlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArg 1260
Db 4208 GCTGGATGGTACACGTTGTTCAGCCAAAGTGAAGCCGGCATCGTGTGTCGACATGCCAGG 4267
Qy 1261 LeuAspIleTrpAlaGlnTrpHisGlnIleProProMetSerValArgProSer 1280
Db 4268 CTGATATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGCTGTCCGCCCCAGT 4327
Qy 1281 GlySerArgTrpGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerSer 1300
Db 4328 GGCAGTCGTCACGGATCTCTCACCAGTAAAGGACTTGACATATTTTCTGCCCTTTTCTCCTC 4387
Qy 1301 MetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAspGluLeu 1320
Db 4388 ATGGAAAGCAGATGGTGTATTTCATGCTCTCTCGGAGTGTAGTGGAGAGTGAATGAACCT 4447

RESULT 2
HSM805275 5804 bp mRNA linear PRI 12-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp451E012 (from clone DKFZp451E012).
DEFINITION AL834247
ACCESSION AL834247
VERSION 1 GI:21739789
KEYWORDS human
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE Wambutt R., Heubner D., Mewes H.W., Weil B. and Wiemann S.
JOURNAL Direct Submission
COMMENT Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
```

consortium of the German Genome Project.
This clone (DKFZp451E012) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mlps.gsf.de/proj/cDNA/>.

FEATURES
Location/Qualifiers
source
1..5804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="DKFZp451E012"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..5804
/gene="DKFZp451E012"
1..4227
/gene="DKFZp451E012"
/note="similarity to actin-associated protein palladin
(Mus musculus)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD38923.1"
/db_xref="GI:21739790"
/translation="YCEKTEGGADKCSGGDRSLSSLOPSSLLKRLTPHSLWTAT
LYQGEFFSPVLSKSPWIIITAVELCDWISSGSTKKKESMODSIEASTSISQ
LURESYLAETHRGNERSRAEPSPNCFHSGAEGGGDDLDLILASLUSBEL
DESVNLARLAIINYPLEKADQARLRSPDQMKHSPNLSFEPFQDNPRTSSKE
SPOEAKRPQVCSSTQSKVFLNKAADIFBELSLFKSHSKRIIPRACKNHKSLKSEA
NKVMQNSSSFDLSRRERSVPIPIPADTRDNEVHALAQEAKRRARAEQAASAA
QKDTTCCSPSSLYVEPLGQPPFTOKLRSREVPEGTRVOLDIVVGPVPPVRYC
EGKELENSPDHIVQAGNLHLTIAEPEDTGRYSCFASNIYCTSTSAEIVIEGVS
SSDEGDPNKEMRNRIQKNEUSVPTTSAVIPAPVPAQHLVAQPRVATIQCCQST
NYLQGLDGPIIAAPVTIKMLQNLSEGLVVEFCRVKGAQSPKVEWREGLIEDS
PDFRIQKKPRMAPEEICTLVIAEFVAFDSCGCTCTASNKYGTVSSLAQLVRGNE
DLRNGSLNASSTTNLAATEPOPPSPHSPVQPPKPKLEGLVLNNEPRSSRI
GUVHNLPEDDKSGSEASAGVVTTRTPDPSFORFENGQATKTPSPSPVKEPPV
LAKPKLDSTQLQQLHNLQVLEQLQNLQNPSPSPKPEFSTVNLNAPVATVTSKQV
KAPSTQFSLARPKYEFPTNTTAAVAFSSPVFTLSSTPQTQITVSKESLVLSPH
SVQTSKPGSTQNEPLPFPTEPTTPTTSPISGNQFQPCVPIVPSVTRIQNP
FAPVLSVLPSPAIPNTAMGLPRSPASMPQSLAKNTKSPQVADDNDIIRTKNAVI
RDLGKKITFSDVRPNOQEVKISSFEQRLMNEIFRLERTPVDESDEIOHDEPTGKC
IAPIDKRLKHFEVTPSPVFTCKTIGVTPVKVWFKDQKQISKNRCKMRREGDG
TCSLHLESTSDDDGNYTMAANPQGRISCSGHLVQSLPIRSRLISAGOSHGRSRY
QERDKPLQERFRFPLQAGDMVAHGRCLRCDCKVSGLPPPELWLLWLLQVLPD
APHKMLVRETGVHSLLDPLTQRDAGTYKCIATNKTGQNSFSLSELSVAKVKKAPVI
LEKLQNGVPEGHPVRLCEVRIGMPVPVFWKDNETIPTCTRERISMHQDTTGYACIL
IOPAKKSDAGWYTLNAGIIVSCATRLDIYAQWHQIIPPPMSVRPSGRSGRYSLTSK
GLDIFSASFSSMESTWYVYSCSSRSRVESDEL"
5766..5771
/gene="DKFZp451E012"
5783
/gene="DKFZp451E012"
BASE COUNT 1650 a 1460 c 1322 g 1372 t
ORIGIN

Alignment Scores:
Pred. No.: 3..55e-252 Length: 5804
Score: 6863.00 Matches: 1313
Percent Similarity: 99.47% Conservative: 0
Best Local Similarity: 99.47% Mismatches: 7
Query Match: 99.61% Indels: 0
DB: 9 Gaps: 0

US-09-818-990b-2 (1-1320) x HSM805275 (1-5804)

Qy 1 MetClnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
|||||
Db 265 ATGCAAGACGACACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGATCAT 324
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40

polyA_signal 5451. .5456
polyA_site 5468
BASE COUNT 1558 a 1392 c 1260 g 1276 t
ORIGIN

Alignment Scores:
Pred. No.: 2,09e-251 Length: 5486
Score: 6842.00 Matches: 1309
Percent Similarity: 99.32% Conservative: 2
Best Local Similarity: 99.17% Mismatches: 9
Query Match: 99.30% Indels: 0
DB: 9 Gaps: 0

US-09-818-990b-2 (1-1320) x HSW803309 (1-5486)

Qy 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 266 ATGCAAGCAGCAGCATAGAAAGCTTCTACCTCCATATCTCAGCTTCTAAGAGAGAGCTAT 325

Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 326 TTAGCTCAACACAGACATCGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCTCCAAC 385

Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu 60
Db 386 CCTTGCCATTTCGGCAGCTCTCTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 445

Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 446 CCAGATCTTCAGCGCTTCTGAGCCAGAGAAGATAGACGAAAGTGCAATTTGGCAAGA 505

Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 506 CTGCCCCATCAATACGACCTTTGGAGAGCGCAGATGAACCTCAAGTAGAAGACGACTT 565

Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 566 TCTCCTGATCAGATGAACACTCACCTAAATTTAAGTTTTCAGCCCTAACCTTCTGCGCAGAT 625

Qy 121 AsnProArgSerProThrSerSerLysLeuSerProGlnGluAlaLysArgProGluTyr 140
Db 626 AACCTCGAAGTCCACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGGCCACCGTAT 685

Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 686 TGTTCTCAACCCAGTCCAAAAAGTATTTTAATTAAGGCTGCGGCTCATTTGAAGAG 745

Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
Db 746 CTATCCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGCTGCAGAAAAAC 805

Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
Db 806 CACAAGAGTAACTGGAAATCTCAAAAAGAGTTATGCGAGGAAACAGCTCCAGTTCTCA 865

Qy 201 AspLeuSerGluArgArgGluArgSerValProIleProAlaAspThrArg 220
Db 866 GATCTGTCAAGAACAGCAGAAAGATCTCTGTTCCCATCCCTATCCCTGCGGATACCAGG 925

Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgGluAlaGluGln 240
Db 926 GATAATGAAGTGAATCACGCCCTGCAACAGCAGGAAGCCGCTGAAGCGGAGCAG 985

Qy 241 AlaAlaSerGluAlaAlaGlyLysPheThrThrProGlySerSerProSerSerLeuTyr 260
Db 986 GCTCCAGTGGAGCGGCTGGTGGAGACACTACACAGGGTCTTCCCTTCTCTCTGATC 1045

Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1046 TATGAAGAACTCTGGGGCAACCTTCCCGGTTCACTCAAAAGTTTACGAGCAGAGAAGTT 1105

Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
Db 1105

Db 1106 CCAGAAGGAATCGAGTACAGTTGGATTGCTAGTAGTGGTCCACCACCTCAAGTA 1165

Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
Db 1166 AGTGGTACTGTGAAGCAAGGAGCTTGAATAATCCCAAGATATTCACATCGTCCAGCA 1225

Qy 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
Db 1226 GGAATCTGCACCTACCTGACCATTTGCGGAAGCCCTTTCAGAGGACACAGAGCGTATTC 1285

Qy 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
Db 1286 TGCTTCTGCTTCTAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1345

Qy 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
Db 1346 GTTCTTCTTCTGACTCAGAGGCGCCCTTAACAGGAAGAGATGAATCCGANTCCAGAG 1405

Qy 381 ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln 400
Db 1406 CCAATGAGGTGTCATCTCTCCACCTACCTCTGCAGCTCATCTCTCCAGCAGTACCCCA 1465

Qy 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
Db 1466 GCCCAGCATTTGTTGGCCCAACCTCGTGTGCAACCATCCAGCAGTGTCAAAGCCCA 1525

Qy 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMet 440
Db 1526 AATTACTTGAGGGATTTGGATGGAAACCTCATCTCAGCTCTCTGTTTACAAAGATG 1585

Qy 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
Db 1586 CTACAAATTTGTCAGCTCTCTGAGGGTCAGCTGGTGTCTTGAATGCAGAGTAAAGGA 1645

Qy 461 AlaProSerProLysValGluTyrPyrArgGluGlyThrLeuIleGluAspSerProAsp 480
Db 1646 GCTCCATCTCTTAAGTTGAGTGTGATAGAGAGGGACTTTAATAGAAGATTTCTCCAGAT 1705

Qy 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu 500
Db 1706 TTTAGGATTTACAGAAAAAACCTTCGATCCATGGCAGCCAGAGGAGATTTGACCTTG 1765

Qy 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
Db 1766 GTCATTGCTGAGGTTTTCAGAGAGGTTCTGGTGTCTTACATGTACTGCAAGCAACAAA 1825

Qy 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
Db 1826 TACGGCACAGTGTCAAGCATTTGCACAGCTGCAGCGTGAGAGAAATGAGGACCTCAGCAAC 1885

Qy 541 AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnPro 560
Db 1886 AACGGGTCTCTTCTACTCAGCCAACTTACCACCAACCTTGGCAGCTATTGAGCCACAGCC 1945

Qy 561 SerProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGly 580
Db 1946 TCCCCACCCCACTCAGAGCTCCATCTCTGGAACAACCCCAACCCCACTCGAGGGG 2005

Qy 581 ValLeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHisPhe 600
Db 2006 GTTCTGGTGAACCAACAATGAGCCCCGGTCCAGCTCCAGGATTTGGCTGCTGTCACATTC 2065

Qy 601 AsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThr 620
Db 2066 GACCTGCTGAAGATGACAAAGGAAGTGAAGCATCTCCCGAGGCTGGTGTGGTCAACCA 2125

Qy 621 ArgGlnThrArgProAspSer**GlnGluArgPheAsnGlyGlnAlaThrLysThrPro 640
Db 2126 AGACAGACCCAGCCCGGATCTTTCAGGAGAGAGTTTCAACGGCAGGACCAACCAACCCCA 2185

Qy 641 GluProSerPheProValLysGluProProProValLeuAlaLysProLysLeuAspSer 660
Db 2186 GAGCCTTCTTCCCCCGTGAAGAGCCCCCTCCAGTTCTGGCCAAACCCCAACCTTGTATTC 2245

```
Qy 661 ThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGlnHisGlnLeuGlnAsnPro 680
Db 2246 ACTCAGTTACAAACAGCTTCATAAACAAGTCTTACTGGAACACACCAATCGCAAAACCCA 2305
Qy 681 ProProSerSerProGlyGluPheProPhe***MetThrValLeuAsnSerAsnAlaPro 700
Db 2306 CTCCTCTCATCTCCCTAAGAGATTCTCTTTCAGCATGACTGTTTTCGAACCTCCAACTCTCC 2365
Qy 701 ProAlaValThrThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeu 720
Db 2366 CACGCGGTACACAACTCCAGTAAGCAGGTGAAGGCTCTTCATCACAGACGTTTCAGCTTG 2425
Qy 721 AlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSer 740
Db 2426 GCCCGCGCGAAGTATTCTTCCCTCCACGACACACCCAGCAGCAACTGTGGCCCTTCC 2485
Qy 741 SerSerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLys 760
Db 2486 AGCTCTCGGGTGTTCACCTTTGAGCAGCAGCTCTCTCAAACTATTTCAGAGACAGTGAGCAAA 2545
Qy 761 GluSerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIle 780
Db 2546 GAAAGCCCTTAGTGTCTACCCCTCTGTGCAAAACCAATCTCCAGGAGGCTTTCCATC 2605
Qy 781 GlnAsnGluProLeuProGlyProThrGluProThrProProProPheThrPheSer 800
Db 2606 CAAAATGAGCCACTCCACAGGCCCAACAGAACCAACACACCAACCACTTCACATTTTCC 2665
Qy 801 IleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThr 820
Db 2666 ATCCCCAGCGGAACACAGTTTCAGCGCCGCTGTGTCTCCCAATTCCTCTCTCTCTACC 2725
Qy 821 SerArgIleGlnAsnProValAlaPheLeuSerValLeuProSerLeuProAlaIle 840
Db 2726 AGCCGGATTACAGAACCCAGTGGCTTCTCAGCTCTGTGTCTGCTCTCTCCCTGCCATC 2785
Qy 841 ProProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeu 860
Db 2786 CCACCCCAAAATGCCATGGGCTGCCTAGAGTGCACCATCCATGCCATCCAGGGATTA 2845
Qy 861 AlaLysLysAsnThrLysSerProGlnProValAsnAspAsnIleArgGluThrLys 880
Db 2846 GCGAAGAAAAATACAAAGTCTCTCAACAGTGAATGATGATGAACATTCGTGAAACTAAG 2905
Qy 881 AsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGln 900
Db 2906 AACGCAGTATTCGAGACTTGGGGAAGAAATAACTTTTCAGTGTGTGACACCAACCCAG 2965
Qy 901 GlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu 920
Db 2966 CAGGAGTACAAAATTTCAAGCCTTGAGCAGAGGCTGATGAATGAATAGATTTCGCTTG 3025
Qy 921 GluArgThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGly 940
Db 3026 GAACGTACTCTCTGTTGATGAATCAGATGATGAATTAACATGATGATGATGATGATGATGAT 3085
Qy 941 LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer 960
Db 3086 AAGTGATTTGCTCCCATTTTGACAAAGAGACTCAAGCACTTCCGGGTGTCACAGAGGCTCT 3145
Qy 961 ProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLys 980
Db 3146 CCAGTCACATTCACCTGCAAAATTTCTGGATACCTGTTCCAAAAGTTTACTGGTTCAAA 3205
Qy 981 AspGlyLysGlnIleSerLysArgAsnGluHisCysLysLysMetArgArgGluGlyAspGly 1000
Db 3206 GATGGGAAGCAGATTTCTTAAGAGAATGAGCACTGCAAAATGAGCGGAGAGAGATGGG 3265
Qy 1001 ThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMet 1020
Db 3266 ACATGCTCTTCACATTTGAATCCACTACCACTGATGATGATGATGATGATGATGATGATGATG 3325
```

```
Qy 1021 AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro 1040
Db 3326 CGAGCAACCCAGGAGGAGAAATCAGCTGTCTGSCCACTTGATGGTACAAAGTTTGCC 3385
Qy 1041 IleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGlu 1060
Db 3386 ATTGCACTCGCTTAACCTCTGCTGGTCAGTCTCACAGGGAAGATCCCGAGTGAAGAA 3445
Qy 1061 ArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGly 1080
Db 3446 AGACAAAGAGCCCTACAGAACGCTTTTCCAGCCACATTTCTCGACGCTCTCTGGG 3505
Qy 1081 AspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuPro 1100
Db 3506 GATATGGTAGCTCATGAGGGCGCTCTGTGCTGGACTGTAAAGTGAGTGGTTTACCG 3565
Qy 1101 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLys 1120
Db 3566 CCCCAGGAGCTGACATGGCTACTCAATGGCAACCTGTGTACCATGACCATCCTCAGCCAA 3625
Qy 1121 MetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAsp 1140
Db 3626 ATGTGGTACAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCCTCAGCGGAC 3685
Qy 1141 AlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGlu 1160
Db 3686 GCAGGGACCTATAAGTGCATCGCTACCAACAAACCCGGGCAAGATTCCTTTAGTCTGGAG 3745
Qy 1161 LeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsn 1180
Db 3746 CTCTCTGTAGTAGCAAGAGGTGAAGAAAGCACTGTGTATCTCTGGAGAAACTACAGAAC 3805
Qy 1181 CysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProPro 1200
Db 3806 TCGCGTGTCTCCAGAGGCCACCCGCTGAGCTGGAGTCCGCGGTGATAGGCGATGCCCCCA 3865
Qy 1201 ProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSer 1220
Db 3866 CCTGTGTCTTACTGGAAGAAAGACAATGAGACCATCCCTTGCACACAGAGAGAGATCAGT 3925
Qy 1221 MetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAsp 1240
Db 3926 ATGCACCAAGGACACAACAGGATGCTCTGCTCTCATTCAGCCAGCCCAAGAAATACAG 3985
Qy 1241 AlaGlyTyrTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArg 1260
Db 3986 CTGATGGTACAGTGTGTCAGCAAGAAATGAGCCGCACTCGTGTGCTGCTGCTGCTGCTG 4045
Qy 1261 LeuAspIleTyrAlaGlnTrpHisGlnIleProProProMetSerValArgProSer 1280
Db 4046 CTGGATATATACGCTCAGTGCACCATCAGATCCACGCGCCCATGCTGTCTCGGCCCATG 4105
Qy 1281 GlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerSer 1300
Db 4106 GGCAGTCTGCTACGATCTCTCACCAGTAAGAGACTTGACATATTTTCTGCTCTCTCTCTCC 4165
Qy 1301 MetGluSerThrMetValTyrSerCysSerSerArgSerValGluSerAspGluLeu 1320
Db 4166 ATGGAAGACCATGGTGTATTTCATGCTCTCTCGAGTGTAGTGGAGATGATGAACCTT 4225
```

```
RESULT 4
HSM803687
LOCUS HSM803687 4683 bp mRNA linear PRI 10-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp451K027 (from clone DKFZp451K027).
ACCESSION AL832379
VERSION AL832379.1 GI:21732942
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4683)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.
```


TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp451K027) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

source

1. .4683

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKFZp451K027"

/tissue_type="cDNA-collection"

/clone_lib="451 (synonym: hlcc1). Vector pSport1; host

DH10B; sites NotI + SalI"

/dev_stage="adult"

polyA_signal

4576. .4581

polyA_site

4594

BASE COUNT 1376 a 1147 c 1048 g 1112 t

ORIGIN

Alignment Scores:

Pred. No.: 9.59e-194 Length: 4683

Score: 5322.00 Matches: 1013

Percent Similarity: 99.41% Conservative: 0

Best Local Similarity: 99.41% Mismatches: 6

Query Match: 77.24% Indels: 0

DB: 9 Gaps: 0

US-09-818-990B-2 (1-1320) x HSM803687 (1-4683)

QY 302 TrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAlaGly 321
DB 293 TGGTACTGTGAAGGCAAGGAGCTTGAATAATCCACATATTCACATCGTCCAGGACGA 352
QY 322 AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys 341
DB 353 AATCTGCACCTACCTGACCATTTGCGAAGCCCTTTGAAGGACACAGACGCTATTCTCTGC 412
QY 342 PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGlyVal 361
DB 413 TTTCTTCTAACATCTATAGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGGGTT 472
QY 362 SerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLysPro 381
DB 473 TCTTCTTCTGACTCAGAAGGCGACCCCTAACAAAGGAGAGATGAATCCAGAACGCA 532
QY 382 AsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGluAla 401
DB 533 AATGAGGTGTCATCTCTCCACTACCTCTGACGATCATCTCTCCAGCAGTACCCCAAGCC 592
QY 402 GlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThrAsn 421
DB 593 CAGCATTTGGTGGCCCAACCTCTGTGGCAACCATCCAGCAGTGTCAAGGCCCAACCAAT 652
QY 422 TyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMetLeu 441
DB 653 TACTTGCAGGGATTGGATGGAAACCTATCATTCAGCTCCTGTTTACAAAGATGCTA 712
QY 442 GlnAsnLeuSerAlaSerGluGlyGlnLeuValPheGluCysArgValLysGlyAla 461
DB 713 CAAATTTGTGACGTTCTGAGGGTTCAGCTGGTTCTTGAATGCAGAGTAAAGGAGCT 772
QY 462 ProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAspPhe 481
DB 773 CCATCTCTAAGGTGTAGTGTAGAGAGGGACTTTAATAGAAGATTTCTCCAGATTTT 832
QY 482 ArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeuVal 501

DB 833 AGGATTTTACAGAAAAACCTCGATCCATGCAGAGCCAGAGGAGATTTTCACCTTGGTC 892
QY 502 IleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyr 521
DB 893 ATTGCTGAGGTGTTTGCAGAAAGATTCTGGGTGCTTACATGTACTGCAAGCAACAATAC 952
QY 522 GlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsnAsn 541
DB 953 GGCACAGTGTCAAGCATTTGCACAGTGCACAGTGCAGAGAAATGAGGACCTCAGCAACAAC 1012
QY 542 GlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnProSer 561
DB 1013 GGGTCTCTTCTACATCAGCCAACTCTACCACCAACCTGTCAGCTATTGAGCCACAGCCCTCC 1072
QY 562 ProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGlyVal 581
DB 1073 CCACCCCACTCAGAGCCTCTCTGTGGAACAACCCCAACCAACCAACCTCAGAGGGGTT 1132
QY 582 LeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsn 601
DB 1133 CTGGTGAACCAATGAGCCCGCTCCAGCTCCAGGATTTGGGCTTCGTGTGCACCTTCAAC 1192
QY 602 LeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThrArg 621
DB 1193 CTGCTCAAGATGACAAAGGAAGTGAAGCATCTCTCCGAGGCTGGTGGTACCACCAAGA 1252
QY 622 GlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu 641
DB 1253 CAGACCAGGCGCGATTTCTTTCCAGGAGAGGTTCAACGACAGGCAACAAACCCACAG 1312
QY 642 ProSerPheProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
DB 1313 CTTTCTTCCCGTGAAGAGCCCTCCAGTCTTGCCCAACCAACCTTGTATTCCTACT 1372
QY 662 GlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGlnGlnHisGlnLeuGlnAsnProPro 681
DB 1373 CAGTTACAACAGCTTCAACCAAGTCTTACTGGAACAACCAACCAATTCGAAAACCCACCT 1432
QY 682 ProSerSerProLysGluPheProPhe***MetThrValLeuAsnSerAsnAlaProPro 701
DB 1433 CTTTCACTCTCTAAGAGGATTTCTCTTTCAGCATGACTGTTTGAACCTCAATGCTCCCA 1492
QY 702 AlaValThrThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeuAla 721
DB 1493 GCGGTGACACATCCAGTAAGCAGTGAAGCTCTCTTCAATCAGACAGCTTCAAGTTGGCC 1552
QY 722 ArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSerSer 741
DB 1553 CGGCCGAAGTATTCTTCTCCCTCCACGAACACCAACCACTGTGGCCCTTCCAGC 1612
QY 742 SerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLysGlu 761
DB 1613 TCTCCGGTGTTCACCTTTGAGCAGCACTCTCAAACTATTCAGAGGACAGTGAAGAA 1672
QY 762 SerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIleGln 781
DB 1673 AGCCTCTTAGTGTCTACCCCTCTGTGCAAAACCAAACTCCAGGAGGCTTTCATCCAA 1732
QY 782 AsnGluProLeuProProGlyProThrGluProThrProProPheThrPheSerIle 801
DB 1733 AATGAGCCACTCCCAAGGCGCAACCAACCAACCACTTCCATTTTCCATC 1792
QY 802 ProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThrSer 821
DB 1793 CCCAGCGAAACCAAGTTTCAGCCCCGTGTGTGCCCAATTCCTGTCTCTCCCTACCAGC 1852
QY 822 ArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIlePro 841
DB 1853 CGGATTCAGAACCCAGTGGCTTCTCTAGCTCTGTTCTGCGCTTCTCTCCCTGCCATCCCA 1912
QY 842 ProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeuAla 861

GPSGAAEGGGQDDLPDLSAFLSQBELDESUNLARLAINYDFLEKADETQARKRLSP
DQMKSPNSFENFQDNPRTSSKESQEAQRPOYQSETQSKKVFNLKAAADFIIE
LSLEKSHSKSRFPNCKNKHKSLEONKMOENSSFSDLSERRERSVPTIPAD
TRNEVNHLEQOEAKRREAQAASEAAGDTPGSSPSLSLYEELGQPPRTQKLR
SRVPEGRVQLQDVIYVPPQVRNYCGBKELENSPDHIHQVQAGNLSLITIAEAE
DTGRYSCFASNIYGTOSTSAEIIYEGVSSDSSEDPKEMNRIQKPNVSSPTTSA
VIPAVFOAHLVAQPRVATIQOCQPTNYLQGLDKPIIAAPVFTKMLQNLASSEGO
LVVFEGRVAGSPKVEWYREGTLIEDSPDFRILQKPRSMABPGKDDFNFLVYE*

BASE COUNT 638 a 501 c 487 g 501 t

Alignment Scores:

Pred. No.: 3, 29e-89 Length: 2127
Score: 2560.00 Matches: 494
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.16% Indels: 0
DB: 9 Gaps: 0

US-09-818-990B-2 (1-1320) x AK027343 (1-2127)

Qy 1 MetcInAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 231 ATGCAAGACGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 290
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 291 TTACTGAACACACATCGGGGAACAATGAGAGGAGTCCGCGGAGCCCTCTCCAC 350
Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu 60
Db 351 CCTTGCCTATTTGGCAGTCTCTTCTGGGGCGCTCAAGGAGCGGAGCCAGATGACCTT 410
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluCluLeuAspGluSerValAsnLeuAlaArg 80
Db 411 CCAGATCTTTCAGCTTCTGAGCCAGAGAAGATTAGACGAAAGTGCAATTTGGCAAGA 470
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 471 CTGCCATCAATTACGACCTTTTGAGAGGCGCAGATGAACTCAAGCTAGAGAACGACTT 530
Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 531 TCTCCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGCTTCTGACCTTCTGCGAGAT 590
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Db 591 AACCTCGAAGTCCACCGCTCTAAAGAAAGCCCGGAGGAGGCAAAAGGCCACAGTAT 650
Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 651 TGTTCGAAACCCAGTCCAAAAGATATTTTAATTAAGGCTGCGACTTCATTGAAGAG 710
Qy 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
Db 711 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGTGCCTGCAAAAAC 770
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSerPheSer 200
Db 771 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTATGCAAGAAAAACAGCTCCAGTTCTCA 830
Qy 201 AspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThrArg 220
Db 831 GATCTGTGAGAAAGACGAGAAGATCTTCTGTTCCATCCCTATCCCTGCGGATACCAAG 890
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
Db 891 GATAATGAAGTGAATCAGCCCTCGAACAGCAGGAGCAAGAGCGCTGAAGCGGAGCAG 950
Qy 241 AlaAlaSerGluAlaAlaGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Db 951 GCTGCCAGTGAGGGGTGGTGGAGACACTACACAGGGTCTTCCCTTCTATCTCTGTAC 1010

Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1011 TATCAACAACCTCTGGGSCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGATT 1070
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
Db 1071 CCAGAAGAACTCGAGTACAGTTGGATTGCTAGTAGGTAATTCACCACTCAAGTA 1130
Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGluAla 320
Db 1131 AGTGGTACTGTGAAGCAAGGAGCTTGAANAATTCACATATTCACATCGTCAGGCA 1190
Qy 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
Db 1191 GGAAATCTGCACCTCACTGACCATTTGCGAAGCCTTTGAAGAGGACACAGAGCGTATTC 1250
Qy 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
Db 1251 TGCCTTCTCTTAACATCTATGGGACAGATTGCGACTTCTGCTGAGATTTATATAGAAGG 1310
Qy 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
Db 1311 GTTCTTCTCTGACTCAGAAGGCGCCCTAACGAAGAGAGATGAATCGAATCCAGAAG 1370
Qy 381 ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln 400
Db 1371 CCAATGAGGTGTCATCTCTCCCACTACCTCTGAGTCATCTCTCCAGCAGTACCCCAA 1430
Qy 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
Db 1431 GCCCAGCATTTGTTGGCCCAACCTCGTGTGGCAACCACTCCAGCAGTGTCCAGAGCCCCACC 1490
Qy 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaProValPheThrLysMet 440
Db 1491 AATTACTTGCAGGATTGGATGGAAAACCTATCATGTGAGCTCTCTGTGTTTACAAAGATG 1550
Qy 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
Db 1551 CTACAAAATTTGTGAGCTTCTGAGGGTCAGCTGGTGTCTTTGAATGCAGAGTAAAGGA 1610
Qy 461 AlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
Db 1611 GCTCCATCTCTTAAGTTGAGTGTAGTAGAAGGAGACTTTAATAGAAGATTCTCCAGAT 1670
Qy 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
Db 1671 TTTAGGATTTTACAGAAAAAACCTCGATCCATGGCAGAGCCA 1712
RESULT 6
AF464873 5773 bp mRNA linear PRI 11-FEB-2002
LOCUS AF464873 Homo sapiens myoneurin (MYN) mRNA, complete cds.
DEFINITION AF464873
ACCESSION AF464873
VERSION AF464873.1 GI:18448934
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5773)
Lockwood, S.K. and Sims, K.B.
TITLE A protein with two immunoglobulin-like domains interacts with the
Norrie disease gene product norrin
JOURNAL Am. J. Hum. Genet. 61, A158 (1997)
REFERENCE 2 (bases 1 to 5773)
Lockwood, S.K.
AUTHORS Direct Submission
TITLE Submitted (31-DEC-2001) Neuroscience Center, Massachusetts General
JOURNAL Hospital, 149 13th Street, Charlestown, MA 02129, USA
FEATURES
Location/Qualifiers
1. .5773
/db_xref="taxon:9606"

Db 1409 TCAATCTCCAAGACAGGGGTGACCACAGCTGTGATT----- 1444
QY 403 HisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThrAsnTyr 422
Db 1445 -----CAACCACTGTCTGCTGTGCAACAGGTTCAACAGTCCAATTCATAT 1492
QY 423 LeuGlnGlyLeuAspGly---LysProIleIleAlaAlaProValPheThrLysMetLeu 441
Db 1493 CTCTGCCGACCTGATGGAACCACTACTGCCTACTTTCCTCTCTTTTACAAAGGAACGTG 1552
QY 442 GlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGlyAla 461
Db 1553 CAAACACACAGCGCTGCGGAAGCCAGGTGGTGTCTGGAGTCCGCGGTCCGTGGGCA 1612
QY 462 ProSerProLysValGluThrPtyrArgGluGlyThrLeuIleGluAspSerProAspPhe 481
Db 1613 CCCCTCTCAGGTCAGTGGTTCGCAAGGAGGTGAATCCAGACTCTCCAGATTTC 1672
QY 482 ArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeuVal 501
Db 1673 CGAATTCACAGAAAACCTAGATACAGCTGACCTGAGGAGATTTCACCCCTAGTT 1732
QY 502 IleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyr 521
Db 1733 ATCGCTGAGACTTTCCTGAAGATGCAGGATCTTTACATGTTTCAGCAAGAAATGATTAT 1792
QY 522 GlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsnAsn 541
Db 1793 GGATCAGCAACAGCACTCCCCCTGGTGTCTACCTCAGCCCAAC-----ACTGAAAAC 1846
QY 542 GlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaAlaIleGluProGlnProSer 561
Db 1847 TGTAGTTACGAGTCAATGGGAGAAATCCACATGACCACTTCCACACTTCCACCTCCC 1906
QY 562 ProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGlyVal 581
Db 1907 CTCCAATCTTGAGACAGTTCCTTGGAGTTGGCTTCAAAGAAACCATCTGAGATCCAG 1966
QY 582 LeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsn 601
Db 1967 CAGGTGAAC---AACCTGAGTTAGCCCTGAGCAGGCGACCCCTCAATGGCAATTCAT 2023
QY 602 LeuProGluAspLysGlySerGluAlaSerSerGluAlaGlyValValThrArg 621
Db 2024 GCTGCTGAGAGGAA-----ACGAGCGAGTCCATCCACGCCGT 2062
QY 622 GlnThrArgProAspSer**GlnGluArgPheAsnGlyGlnAlaThrLysThr----- 639
Db 2063 -----GGAGTAAATGGACTGATTAAACGGCAAGCTAACAGTAATAATCT 2107
QY 640 ---ProGluProSerPhe-----ProValLysGluProProValLeuAlaLys 655
Db 2108 CTTCACACACGACTGCTCGCTTCCACCCACTAAAGAGCCACCACTCTGCTGGCCAAA 2167
QY 656 ProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGlnHis 675
Db 2168 CCAAACTA----- 2176
QY 676 GlnLeuGlnAsnProProProSerSerProLysGluPheProPhe**MetThrValLeu 695
Db 2176 ----- 2176
QY 696 AsnSerAsnAlaProProAlaValThrThrSer**LysGlnValLysAlaProSerSer 715
Db 2176 ----- 2176
QY 716 GlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAla 735
Db 2176 ----- 2176
QY 736 ThrValAlaProSerSerSerProValPheThrLeuSerSerThrProGlnThrIleGln 755

Db 2176 ----- 2176
QY 756 ArgThrValSerLysGluSerLeuValSerHisProSerValGlnThrLysSerPro 775
Db 2176 ----- 2176
QY 776 GlyGlyLeuSerIleGlnAsnGluProLeuProProGlyProThrGluProThrProPro 795
Db 2176 ----- 2176
QY 796 ProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerProIle 815
Db 2176 ----- 2176
QY 816 ProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuPro 835
Db 2176 ----- 2176
QY 836 SerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMet 855
Db 2176 ----- 2176
QY 856 ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAspAsn 875
Db 2177 -----GGATTTCCAAAGAGGCCAGTAGAACTGCTAGAAATAGCCTCCGATGAGAA 2227
QY 876 IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAsp 895
Db 2228 ATTCAGGACCAAGAGTGTGTATTCAAGACCTGGAAGCAAACTTCGCTTCAAGGAG 2287
QY 896 ValArgProAsn----- 899
Db 2288 GACCTCTGACAACTGGCCAGCCGAGTTAAACATACGAAGAAAGATGGCTCGTGACTG 2347
QY 900 -----Gln 900
Db 2348 CTAGGTGCTGACAGTGCAACTGTCTTAAATATTTCAGGAGCCAGAAGAGAAACAGCTAAT 2407
QY 901 GlnGluThrLysLysSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu 920
Db 2408 CAGGAATACAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAGAGTACAGGCTA 2467
QY 921 GluArgThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGly 940
Db 2468 GAAAGTCTCTGTGGATGAATCAGGTGATGAAGTTCAGTATGGAGATGTCCTGTGGAA 2527
QY 941 LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer 960
Db 2528 AATGGAATGGCACTTCTTTGAGATGAAGCTGAAACATTTACAAGATCTTTGAGGGAATG 2587
QY 961 ProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLys 980
Db 2588 CCAGTAACCTTACATGTAGATGGCTGGAAATCCAAAGCCAAAGATCTATTGTTTAAA 2647
QY 981 AspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGly 1000
Db 2648 GATGGGAAGCAGATCTCTCCAAAGAGTGATCACTACACCATTTCAAAGAGATCTCGATGG 2707
QY 1001 ThrCysSerLeuHisIleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMet 1020
Db 2708 ACCTGCTCCCTCCATCCACAGGCTCCACCTAGATGATGATGGAAATTTATACAAATTTATG 2767
QY 1021 AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro 1040
Db 2768 GCTGCAAAACCTCAGGCGCGCATGTTGTACTGGAGCGCTAATGTGTACAGGCTGTCAAC 2827
QY 1041 IleArgSerArgLeuThrSerAlaGlyGlnSerHis-----ArgGly 1054
Db 2828 CAAAGAGTCCAGTCCCCGCTCTCCCTCAGGCCATCTCATGTCAGAAAGCCCTCGTCT 2887
QY 1055 ArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHis 1074
Db 2888 AGATCAAGGACACATGGAGACGAAAAATGAACCAATTTTCAGGAGCGATTCTTTCAGACCTCAC 2947

Qy 1075 PheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCys 1094
 Db 2948 TTCTTGCAGGCTCTGGAGATCTGACTTTCAAGAAGAAACTCTGCAGAAATGGACTGC 3007
 Qy 1095 LysValSerGlyLeuProProGlyLeuThrTrpLeuLeuAsnGlyGlnProValLeu 1114
 Db 3008 AAAGTCAGTGGTTACCAACCCAGATCTAAGCTGGCAACTAGATGGAAGACCCGTCAGC 3067
 Qy 1115 ProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAsp 1134
 Db 3068 CCTACAGTGTCTACAGATGCTGGTGGTGAGAACGGGTGCACCTCTCTGATCATAGAG 3127
 Qy 1135 ProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGln 1154
 Db 3128 CCAGTCAGCTCACGTGATCGCGCATCTACACATGTATAGTACCACCGAGCAGCAG 3187
 Qy 1155 AsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaProValIle 1174
 Db 3188 AACTCATTTACGCTGGAGCTTGTGTTGCTGCTAAAGAACACACAAACCCCTGTGTTT 3247
 Qy 1175 LeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArg 1194
 Db 3248 ATTGAGAACTCCAAACACAGAGAGTGTCTGATGGGTACCCAGTGCAGCTGGAATGTCGT 3307
 Qy 1195 ValIleGlyMetProProValPheTrpTrpLysLysAspAsnGluThrIleProCys 1214
 Db 3308 GTATTGGAGTGCCACCACTCAGATATTTTGGAGAAAGAAATGAATCACTCACTCAC 3367
 Qy 1215 ThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGln 1234
 Db 3368 AGCACTGACCGAGTGAGCATGCACAGCAGCAACACGCGCTACATCTGCTGCTCATTCAG 3427
 Qy 1235 ProAlaLysLysSerAspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIle 1254
 Db 3428 GGAGCCACAAAGAGATGCTGGTGGTGATGCTGTCAGCCAAAGATGAAGCAGGGATT 3487
 Qy 1255 ValSerCysThrAlaArgLeuAspIleTyr-----GlnIleProProMetSerVa 1277
 Db 3548 GGAGAACAAATACCAACCCAGTGGCATCAGCATGCACAGCAGCAACCAAAAAAGT 3607
 Qy 1277 lArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAl 1297
 Db 3608 ACGCCCTCACCGAGTCGCTATGAGCAGCTTTCGGACCAAGGACTAGACATCAAGCAGC 3667
 Qy 1297 aPheSerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSe 1317
 Db 3668 GTTCCAACT---GAGGCCAACCCATCTCACCTGCACACTGAATACTGCTTGTGTAGAAAG 3724
 Qy 1317 rAspGluLeu 1320
 Db 3725 TGAGGACCTG 3734

RESULT 7
 AK095512 2920 bp mRNA linear PRI 15-JUL-2002
 LOCUS Homo sapiens cDNA FLJ38193 fis, clone FCBBF1000280.
 DEFINITION AK095512
 ACCESSION AK095512
 VERSION AK095512.1 GI:21754783
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens fetal brain cDNA to mRNA, clone_lib:FCBBF1
 clone:FCBBF1000280.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
 Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
 Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A.,

Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K.,
 Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y.,
 Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K.,
 Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,
 Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
 and Isogai,T.
 NEDO human cDNA sequencing project
 Unpublished
 REFERENCE 2 (bases 1 to 2920)
 AUTHORS Isogai,T. and Yamamoto,J.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.
 FEATURES
 source Location/Qualifiers
 1. 2920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="FCBBF1000280"
 /tissue_type="brain"
 /clone_lib="FCBBF1"
 /dev_stage="fetal"
 /note="cloning vector: pME18SFL3"
 BASE COUNT 839 a 736 c 718 g 627 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.99e-71 Length: 2920
 Score: 2081.50 Matches: 475
 Percent Similarity: 51.79% Conservative: 132
 Best Local Similarity: 40.53% Mismatches: 249
 Query Match: 30.21% Indels: 318
 DB: 9 Gaps: 23

US-09-818-990b-2 (1-1320) x AK095512 (1-2920)

Qy 228 LeuGluGlnGlnGlnAlaLysArgArgGluAlaGluGlnAlaLaserGluAlaGly 247
 Db 13 CTGCGCCGCCAGGAGGTCTCCCGGAGCGCGCGGCCAGCCTCTCGAGCGCAGCAG 72
 Qy 248 GlyAspThrThrPro---GlySerSerProSerSerLeuTyrTyrGluGluProLeu--- 265
 Db 73 GCAAGGGCGCGCGCTTTTGGCAGCGCGCTGCCAGTTAAATTT-----TCTCCGTTGCAT 126
 Qy 266 -----GlyGlnProProArgPheThrGlnLysLeuArgSer 277
 Db 127 TCTGCAACAGAGATCGCATTCAGAGCCCAAGATTGCGAAGTGGGAATAAGCG-AGATCT 185
 Qy 278 ArgGluValProGluGlyThrArgValGlnLeuAspCysIleValValGly----- 294
 Db 186 GAAAGACCCCAAGAGAGCT-----TGTGGAGGTGGGGCAGACGGC 230
 Qy 295 ---lIleProProGlnValAlaArgTyrTyrCysGluGlyLysGluLeuGluAsnSerPro 313
 Db 231 GTTTTTCGCCCA-----ACTGA-TGTTTCTGTGAAGGAAAGAACTGCACACACTCCT 283
 Qy 314 AspIleHisIle---ValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPhe 332
 Db 284 GATATTCAATCCACTCTGAGGGAGGGGACCTCCATACCTGATCATAGCAGAGCCGCTTT 343
 Qy 333 GluGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThr 352
 Db 344 GAGGACACACAGCTCGCTACACCTGTGTTGCTACGAATCCACCGGCTCAGACACACA 403
 Qy 353 SerAlaGluIleTyrIleGluGlyValSerSerSerGluGlyAsp----- 369

Db 404 TCTGCTGAGGTGTTATGAAGTGCAGTTCACAGATTCTGACAGTGAAGTTAGCT 463
QY 370 -----ProAsnLysGluGluMetAsnArgIleGlnLysProAsnGlu----- 383
Db 464 TTCAAATCAAGAGCTGGAGTTCGCCACAGCTCAAAAGAAACAACACTTCTGTTCTCTTG 523
QY 384 -----ValSerSerProProThr-----ThrSerAlaValIleProProAlaVal 398
Db 524 ACAATAGGATCATCATCCAAAGACAGGGGTGACCACAGCTCTGATT----- 571
QY 399 ProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSer 418
Db 572 -----CAACCACCTGCTGCTCCCTGTGCAACAGGTTCACAGT 607
QY 419 ProThrAsnTyrLeuGlnGlyLeuAspGly---LysProIleIleAlaAlaProValPhe 437
Db 608 CCAACTTCATATCTGCGGACCTGATGGAAACCACCTACTGCTCTTCTCCTGTTTT 667
QY 438 ThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArg 457
Db 668 ACAAGGAACCTGCAAAACACAGCCGTCGGGAAGCCAGGTGTCTGAGTGCCGG 727
QY 458 ValLysGlyAlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAsp 477
Db 728 GTCGGTGGGCGACCCCTCTGTCAGGTCCAGTGTGTCGCAAGGAGTGAATCCCAAGAC 787
QY 478 SerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIle 497
Db 788 TCCTCCAGATTCCGAATCTACAGAAAAACCTAGATCTACAGCTGAACCTGAGGAGATT 847
QY 498 CysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAla 517
Db 848 TGCACCTAGTTATCCGGAGACTTCCCTGAAGATGCCAGGATCTTACATGTTAGCA 907
QY 518 SerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAsp 537
Db 908 AGAAATGATTATGGATCAGCAGCAGCAGTGCAGCTGCTGTCACCTCAGCCAAAC--- 964
QY 538 LeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaAlaIleGlu 557
Db 965 ---ACGAAACCTGTATGACAGTCAATGGGAGATCCAAACATGACCCTCCCAACAC 1021
QY 558 ProGlnProSerProProHisSerGluProProSerValGluGlnProProLysProLys 577
Db 1022 TTTCCACCTCCCTCCCAATCTTGGAGACAAGTTCCTTGGAGTTGGCTTCAAGAAACCA 1081
QY 578 LeuGluGlyValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArg 597
Db 1082 TCTGAGATCCAGCAGGTGAAC---AACCCCTGAGTTAGGCTTGACGAGGCGCCCTTCAA 1138
QY 598 ValHisPheAsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyVal 617
Db 1139 ATGCAATTCATGCTGCTGAGAGGAAACCAAC---GGAGTC 1177
QY 618 ValThrThrArgGlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThr 637
Db 1178 CATCCCGAGCGT-----GGAGTAAATGGAGTATTAACGGCAAGCTTAAC 1222
QY 638 LysThr-----ProGluProSerPhe-----ProValLysGluProProPro 651
Db 1223 AGTAATAAATCTCTTCAACACACAGCTGCTGCTTTCACCCCACTAAGGAGCCACCACT 1282
QY 652 ValLeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeu 671
Db 1283 CTGCTGCCAAACCAAACTA----- 1303
QY 672 LeuGluGlnHisGlnLeuGlnAsnProProProSerProLysGluPheProPhe*** 691
Db 1303 ----- 1303
QY 692 MetThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLys 711

Db 1303 ----- 1303
QY 712 AlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsn 731
Db 1303 ----- 1303
QY 732 ThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSerSerThrPro 751
Db 1303 ----- 1303
QY 752 GlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerValGln 771
Db 1303 ----- 1303
QY 772 ThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyProThrGlu 791
Db 1303 ----- 1303
QY 792 ProThrProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCys 811
Db 1303 ----- 1303
QY 812 ValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSer 831
Db 1303 ----- 1303
QY 832 SerValLeuProSerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSer 851
Db 1303 ----- 1303
QY 852 AlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProVal 871
Db 1304 -----GGATTTCAAAGAGGCCAGTACAGTCTAGTAATGACC 1342
QY 872 AsnAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIle 891
Db 1343 TCCGATGAGGAAATTCAGGCACAAAGAGTCTGTTATTCAAGACCTGGAAACGAAACTT 1402
QY 892 ThrPheSerAspValArgProAsn----- 899
Db 1403 CGCTTCAAGGAGGAGCTCTCTCAACAATGGCCGCGAGTTACATACGAAGAAAGATG 1462
QY 899 ----- 899
Db 1463 GCTCGTGCAGTGTAGTGTGACAGTGCACATGCTCTTTAATATTACGAGGCCAGAGAG 1522
QY 900 -----GlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIle 916
Db 1523 GAAACAGCTAATCAGGAATACAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATA 1582
QY 917 GluPheArgLeuGluArgThrProValAspLysSerAspAspGluIleGlnHisAspGlu 936
Db 1583 GAGTACAGGCTAGAAAGCTCTCTGTGATGAATCAGGTGATGAAGTTCAGTATGGAGAT 1642
QY 937 IleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgVal 956
Db 1643 GTGCTGTGGAATAATGGAATGGCCACCATCTTTGAGATGAAGTGAACATTTACAAGATC 1702
QY 957 ThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysVal 976
Db 1703 TTTGAGGAATGCCAGTAACCTTTCCATGATGAGTGGCTGGAATCCAAAGCCCAAGATC 1762
QY 977 TyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArg 996
Db 1763 TATTGGTTTAAAGATGGGAAGCAGATCTCTCCAAAGAGTATGATCACTACACATTTCAAAGA 1822
QY 997 GluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsn 1016
Db 1823 GATCTCAGTGGACCTGCTCCCTCCATACACACCTCCACCTTAGATGATGATGGAAT 1882
QY 1017 TyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetVal 1036
Db 1883 TATACAATTATGGTGCAAACCCCTCAGGGCGCATCAGTCTGCTACTGGACGCTAATGGTA 1942

Db 527 TCAGCTCCGAAACAGGAGTGACACAGCTGTGTGATTGAGCCCTATCTGTGCTGTCCAA 586
Qy 415 GlnCysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle---IleAla 433
Db 587 CAGGCTCACAGCGCTACTTTCATATCTCTGGGACCTGACGGAACCCACCATGGGCTGCTT 646
Qy 434 AlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValVal 453
Db 647 CTTCTCTGTTTTACTAAGGAACCTACAGAACACAGCAGCCTCCGAGGCGCAGGTGGTGGT 706
Qy 454 PheGluCysArgValLysGlyAlaProSerProLysValGluTrpTyrArgGluGlyThr 473
Db 707 CTGAGTGCAGAGCCGAGGGGACCCCGCTGACGATCCAGTGGTTCCGCGAAGGAAGT 766
Qy 474 LeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGlu 493
Db 767 GAGATCCAGAGCTCTCCGGACTTCAGAAATCTCCAGAAAACCCAGATCGACAGCTGAA 826
Qy 494 ProGluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 827 CCCGAGGAGATCTGACCTTGTCATCTGAGAGTTTCCCTGAAGATCGAGGCATCTT 886
Qy 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
Db 887 ACCTGCTCAGCCACAAATGACTACGGCTCAGTGACACGACCTGACACAGCTTGTTATA--- 943
Qy 534 GlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeu 553
Db 944 -----ACCTCCGCCAACAGAGAACTGTAGTTACGACTCAACAGGAGAACCCACAGT 997
Qy 554 AlaAlaIleGluProGlnProSerProPro----- 563
Db 998 GATCATTCCAACTTTCCACTTCCACTCCCTCCGATCTGGAGACAGCTCCTATGAGCTG 1057
Qy 564 ---HisSerGluProSerValGluGlnProProLysProLysLeuGluGlyValLeu 582
Db 1058 GCATCCCAGAACCATCTGAAATCCAGCAGGTGAACAGCCCAATTTA----- 1105
Qy 583 ValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHisPheAsnLeu 602
Db 1106 -----GGATTTAGCATGGCAGCTCTTCAAAATGCAATTCAACACT 1144
Qy 603 ProGluAspAspLysGlySerGluAlaSerSerGluAla-----GlyValValThrThr 620
Db 1145 GCAGAGGAGGAGACCAACGGAGTCCATCCAGCCTGAGTCAACGGCGCTGATT----- 1198
Qy 621 ArgGlnThrArgProAspSer**GlnGluArgPheAsnGlyGlnAla-----Thr 637
Db 1199 -----AATGCAAGCTTATGCAATAAA 1222
Qy 638 LysThrProGluProSer-----PheProValLysGluProProProValLeuAla 654
Db 1223 TCTCTCCAACACCACTGCGCTGTTACCCACTTAGGAACCCACCCCACTCCTTGCC 1282
Qy 655 LysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGln 674
Db 1283 AAACCAAACTGGATCTCTGAAGCTTCAGCAGCTGCAGAACCAAGTGCCTGGAACAG 1342
Qy 675 HisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe**MetThrVal 694
Db 1343 GAGGCTGCGGCTGCGCCCA-----GCGCCCGCAGGTGTCCCTTGCAACAGCAGCAGC 1396
Qy 695 LeuAsnSerAsnAlaProProAla----- 702
Db 1397 AGCGGCAGCAGCGCGCCCGCTGCGCGCCCTTCCCCCACCAGCCACCCCTTTCGGAG 1456
Qy 703 -----ValThrThrSer***LysGln 709
Db 1457 CTCGGGCGCTGCGTCCCGGTGCCCTCGGAGCCCATGAGCGCTCTCGGCTCCCGCGCT 1516
Qy 710 ValLysAlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSer 729
Db 1517 ACCGCCATGCACTCTCGGCTCTTCAACTACGCGCGCCCTTAAGCAGTTTCATCGCGCGC 1576

Qy 730 ThrAsnThrThrAlaAlaThr-----ValAlaProSerSerSerProValPheThrLeu 747
Db 1577 CAGAACCTGGGTCCCGATCGGGCTGCCACGCAACTCCAGCCCG-----AGTCC 1630
Qy 748 SerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHis 767
Db 1631 TCCAGCTTGGCTCACCATTGTCGCCAGCCAGGCCCTTCGCGCGCGCGCGCGCG 1690
Qy 768 ProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProPro 787
Db 1691 CCCTTCGTGGAACCCGAGGCCATGTGGGCGCGCTCTCGCCTCGCGCCACCCGCCCA 1750
Qy 788 GlyProThr-----GluPro 792
Db 1751 CCTCGGTCCTCAGCCCTCTGACGCTACCCGGTGGCCGAGCTGTTCCTCCACTGCCACG 1810
Qy 793 ThrProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysVal 812
Db 1811 CCGCGCGCACCG-----CTAGCTAGCTCCACTCA-----CACTGCGCC 1849
Qy 813 SerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSerSer 832
Db 1850 TCGCCCGCGCTTCGCGCCCGCAG-----CAGACGCCCGCAGCCTTCTCAGCGCC 1900
Qy 833 ValLeuProSerLeuProAlaIleProProThrAsnAlaMet**LeuProArgSerAla 852
Db 1901 CTGCTGCCCTCTCAGCCACCGCCTGTGCTGTCAACGCCCTGGGGCTGCCCAAG----- 1954
Qy 853 ProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsn 872
Db 1955 --GGAGTCAACCCCGCGGGTTCCCAAGAAGTCCAGTAGAAGTCTAGAAATTCCTCT 2011
Qy 873 AspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThr 892
Db 2012 GATGAGGAGATTCAAGGCACAAAGATGCTGTCTCAAGACCTTGAACGGAAGCTTCGC 2071
Qy 893 PheSerAspValArgProAsn----- 899
Db 2072 TTCAGGAGGAGCTTCTGAACAATGGCCACCGAGGCTAACTATGAGGAAGAATGCT 2131
Qy 899 ----- 899
Db 2132 CGCGCCTGCTGGAGCGCAGCGCAACGCTCTTCAACATCCAGGAGCCAGAGAAACG 2191
Qy 900 -----GlnGlnGluTyrLysSerSerPheGlnGlnArgLeuMetAsnGluIleGlu 917
Db 2192 GCAGCCCAATCAGGAGTACAAAGTCTCTAGCTGCGAGCAGAGGCTGATTAGCGAGATTGAG 2251
Qy 918 PheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAspGluIle 937
Db 2252 TACCGCTGGAGCGCTCCCTGTGGATGAGTCCGGGAGACGAGGTGCAGGATCCAGATGTG 2311
Qy 938 ProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThr 957
Db 2312 CTTGTGGGAACGCAACAGCTCCCTTCTTGTAGATGAAGCTGAAACACTACAAGATCTTT 2371
Qy 958 GluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyr 977
Db 2372 GAGGGGATGCGCGTGACTTTTCCAGTGTGAGTGGGTAATCCAAAGCCAAAGATCTAT 2431
Qy 978 TrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGlu 997
Db 2432 TGGTTTAAAGATGGGAACAGCATTTCTCCGAGAGCGGATCACTACACCATTCAGAGAC 2491
Qy 998 GlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyr 1017
Db 2492 CTGATGGGACTGCTCTCTCCACACCGCGCTCTACCTAGACGACGATGGGAATAC 2551
Qy 1018 ThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln 1037
Db 2552 ACCATCATGGTGCACACCCCTCAGGGTGCCTGCTAGTTGTACAGGAAGGCTATGTGATAG 2611


```
QY 1038 SerLeuProIleArgSerArg-----LeuThrSerAlaGlyGlnSerHisArgGlyArg 1055
Db 2612 GCTGTCAACAAAGAGCGCGAGTCCCGCTCCCTCAGGCGCATCTCCAGCAAGG 2671
QY 1056 SerArgValGlnGluArgAsp-----LysGluProLeuGlnGluArgPhePhe 1071
Db 2672 CCTCGCTCTCGATCAGCGGAGCAGTGGAGATGAACACGACCCATTTCAGGAGCGATTCTTC 2731
QY 1072 ArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArg 1091
Db 2732 AGACCTCACTTCTCGAGCTCTCGAGACCTGACCGTTCAGGAAGGCAAGCTCTGCAGG 2791
QY 1092 LeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGln 1111
Db 2792 ATGGACTGCAAGTCAAGTGGTGTACCAACCCAGATCTCAGTGGCAACTAGATGGAAG 2851
QY 1112 ProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1131
Db 2852 CCCATACCCCGGACAGTGTCTACAAAGATGCTGTGCTGAGATGGGTGCCACTCCCTC 2911
QY 1132 LeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLys 1151
Db 2912 ATTAGAGCCAGTCACTCCCGGAGCGCGCATCTACACATGTATTGCCACCAACAG 2971
QY 1152 ThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAla 1171
Db 2972 GCAGGACAGAACTCGTTTAACTGGAGCTGTGTGCTGTCTAAGGAAGCACACAGGCC 3031
QY 1172 ProValIleLeuGluLysLysGlnAsnCysGlyValProGluGlyHisProValArgLeu 1191
Db 3032 CCTGTGTTTATGGAGAAGCTACAGAACACGGGGGTGTGTGATGATACCCAGTCCGCTG 3091
QY 1192 GluCysArgValIleGlyMetProProValPheTyrTrpLysLys 1207
Db 3092 GAATCCGCTGTCTCGGAGTCCGCCACCTCAGATATTTTGNAAAAA 3139

RESULT 9
AC125500/c
LOCUS
DEFINITION AC125500 136493 bp DNA linear HTG 23-AUG-2002
pieces
Takifugu rubripes clone 206P3, WORKING DRAFT SEQUENCE, 2 ordered
pieces
AC125500
AC125500.2 GI:22450574
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
Takifugu rubripes
SOURCE
Takifugu rubripes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 136493)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Hagnigni,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
McGulles,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paquirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
Stantipop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wecherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NTSC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 136493)
Green,E.D.
Direct Submission
Submitted (28-JUN-2002) NTH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 136493)
Green,E.D.
Direct Submission
Submitted (23-AUG-2002) NTH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
```

COMMENT

On Aug 23, 2002 this sequence version replaced gi:21622707.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@hgrl.nih.gov
----- Project Information
Center project name: dif
Center clone name: 206P03

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 136144 bases at least Q40
Consensus quality: 136285 bases at least Q30
Consensus quality: 136382 bases at least Q20
Insert size: 130000; agarose-fp
Quality coverage: 11.03x in Q20 bases; agarose-fp
Quality coverage: 10.51x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 135196: contig of 135196 bp in length
* 135197 135296: gap of unknown length
* 135297 136493: contig of 1197 bp in length.

FEATURES

source
1..136493
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="206P3"
/clone.lib="Incyte Genomics"
misc_feature
1..135196
/note="assembly_fragment
clone_end:T7
vector_side:left"
1..11443
/note="clone overlaps with GenBank Accession Number
AC124911 clone 209H6 (center project name die)"
misc_feature
135297..136493
/note="assembly_fragment
clone_end:SP6
vector_side:right"
BASE COUNT 38625 a 29019 c 29646 g 39103 t 100 others
ORIGIN

Alignment Scores:

Pred. No.: 3,53e-64 Length: 136493
Score: 1953.50 Matches: 779
Percent Similarity: 24.00% Conservative: 173
Best Local Similarity: 19.64% Mismatches: 304
Query Match: 28.35% Indels: 2716
DB: 2 Gaps: 66


```
US-09-818-990b-2 (1-1320) x AC125500 (1-136493)
QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 85826 ATCAGAGGACACA---GAGCAGCTCAGTCTCCTGTCTCAGCTCTTAAGGGAGGCTAC 85770
QY 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 85769 ATGGCTGAGGCC-----AGGGCCAGCATCGTCACAGCGCGCAGATGCCCTCCTCTCG 85716
QY 41 ProCysHisPheGlySerPro---SerGlyAlaAlaGluGlyGlyGluGlnAspAsp 59
Db 85715 CACCTTCAGTTCACCTGCCCTTGAAGGGTAAAGCAATGACCACATGGGCCACGGCAT 85656
QY 60 Leu-----ProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGluSerValAsn 77
Db 85655 GTCCATTTCCAGATCTCTCAGCCCTTCCTGAGCCAGAGAGAGCTGGACAGAGTAAAC 85596
QY 78 LeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArg 97
Db 85595 CTGGCCCGGAGCCCATCGGACAGAG-----CACAGGGCAGAGGTCAGC 85551
QY 98 LysArgLeuSerProAspGlnMetLys-----HisSerProAsnLeuSerPheGluPro 115
Db 85550 GCCTCTGTACCCCGCAGCAGCCCTGTCCAAACCCCACTCATCTCCTCAGCCTCATCTGAACC 85491
QY 115 ----- 115
Db 85490 CTGCCTGTCCATTGGCTAACCACTCAGTGGGCCCCCTTACC GCCCCCAACACAGCCA 85431
QY 116 ----- 116
Db 85430 CTTCTCAATTCAACAACTGCCAGCAGCTCACATATCTCTTACCCCAAGACAGCAGATG 85371
QY 124 SerProThrSerSerLys----- 129
Db 85370 CACAAGACATCTTCAGGGCATGATGACATGATCAGGAACAGCGGGACAGCGCGAAGGA 85311
QY 130 -----GluSerProGlnGluAlaLysArgProGluTyrCysSerGluThrGlnLys 147
Db 85310 ATCAATGATCCCAACAGGGGGGGAGAACGCTCCCTATTAACCTGGAGACCCAGTCCAAG 85251
QY 148 LysValPheLeuAsnLysAlaAlaAspPheIleGluLeuSerSerLeuPheLysSer 167
Db 85250 AAGAATTCTCAACAAGCAGCGGACTTTATCGAGGAGCTCTTCTCTGTTCAGGCT 85191
QY 168 HisSerSerLysArgIleArgProArgAlaCysLysAsnHisLysSerLysLeuGluSer 187
Db 85190 AACAGCTCCAAACGGATACGACCCAGATCTTGCAATCCCACTCCACCGGACTAGG-----AGC 85137
QY 188 GlnAsnLysValMetGlnGluAsnSerSerPheSerAspLeuSerGlu---ArgArg 206
Db 85136 CAGAACAAG-----AGCCACAATGATGGATCATTTGATCCCTGGATGAGGATACCAGG 85083
QY 207 GluArgSerSerValProIle-----ProIleProAlaAspThrArg 220
Db 85082 GAGCGTGCCCTCATGTCACAGAGGAGGAGGAGGAGGCGCTTACCCTGCT----- 85032
QY 221 AspAsnGluValAsnHisAlaLeuGluGlnGln----- 231
Db 85031 -----GTTGCCCAACCAAGAGAGGAGGAGCGCGGACGCTGGGATCGGCCACTTTCAG 84981
QY 232 -----GluAlaLysArgArgGluAla-----GluGlnAlaAlaSerGlu 244
Db 84980 GACTGCAGGATTACTAGGCGCTGCCAGAAAACCTGTCTGCAGCAGGACGCTCGAGAAGAG 84921
QY 245 AlaAlaGlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGluGluPro 264
Db 84920 ACCGAGGTCGTCTCTGACAGAGAGCTCTACCCAGCA-----GAGCCT 84876
QY 265 LeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGlyThr 284
Db 84876 ----- 284
```

```
Db 84875 GTATGTGAGCGCGCTCATTTTATCCAGAAACTGAAGAGCAGGAGATCGCCGAGGCAGC 84816
QY 285 ArgValGlnLeuAspCysIleValValGlyIleProProGlnVal----- 300
Db 84815 AAGTGCAGTGGACATCGTAAAGAGGATCCCGCTGCTGAAGTCAGTGAAGCAAC 84756
QY 300 ----- 300
Db 84755 CCAACAGACACCAGGCTGACCTGTGACTTCTTAATCATGTTCTCTCAAAACGAGCCGAAC 84696
QY 300 ----- 300
Db 84695 CTCGAGGATGTTTGTCTCCACGGATGACGTACAGATCTACAGAGTGGAGCTGTGGGTCA 84636
QY 300 ----- 300
Db 84635 CGTTTCATTCAAACTCACTGAATCTGAAGTAAATATTTCATATTTAAACCATTTATATTA 84576
QY 300 ----- 300
Db 84575 TATCATCCACTTAACATTCGGTTTAGTTCGGTATTAGTTGTTTAGTAGAGTTTTTCAA 84516
QY 300 ----- 300
Db 84515 GAAGTTGTTGGTCAAGCTTTTCATCGGCCAGACTTGTAGTAAATGTATGAAGTGTGGC 84456
QY 300 ----- 300
Db 84455 ATGAGATAAACCTAAAGTGAACATTGTGAAATCAGCTGTTGTTTCAATAGTGATTTC 84396
QY 300 ----- 300
Db 84395 ACACAAGTGAAGAAGATGATTAATCAGAGGCCAAGAAATTGAACACCTGGAGACATCAC 84336
QY 300 ----- 300
Db 84335 ATGAGCAGAGCTGCGATGATGCTGAAGTTCCTCAGTCATGATGCTGTTTTCGATACTT 84276
QY 300 ----- 300
Db 84275 GCAGACAAAGTGACATGTGATGACAGCTGTTCTGAAACGGTGCATGCTGATTAACAG 84216
QY 300 ----- 300
Db 84215 AACCCCGAGGTTCCCTGCTTCATAGAATGATGTCACAAACACGACCCACATGT 84156
QY 300 ----- 300
Db 84155 TGACCCCTCGGACAGAAATGAAACATTTTTCAGCAAGTGAGATGGATTTAGCCATCGTT 84096
QY 300 ----- 300
Db 84095 CTGGCTATTTCGTGACACTTATGTTAACCTTGCTGTAGTCACAGGGTGTACAGACTATA 84036
QY 300 ----- 300
Db 84035 GGGGCGATGGCTGTGGTCTCACAACATTTGTAGGCCATGAACGTACACATTTGTCTGG 83976
QY 300 ----- 300
Db 83975 AGAATGTCGCTCTCTGCCATTTTAGCCCGGACACAGCAGAGCTGTAGTTGTGTC 83916
QY 300 ----- 300
Db 83915 TGATAATGGAACCGAGGAGCTCTTGACAGCAGCGTGACGTGTGTAAGGTGCTTTTAC 83856
QY 300 ----- 300
Db 83855 ACTGACATTCCTCAGCCAGTATCTTTATTTTCTCATTTTAGAGAGCTGCTTTTGTCTTTT 83796
QY 300 ----- 300
Db 83795 TCTGCATGTCCCCAAATGTGCCACTAGTTTGTGGTGGGTTTGCATTTGTGCCACGA 83736
```

QY 300 ----- 300
Db 83735 ATGTTTATGATGTTTAAACCTAAACACTGCCGCTGTCATCACTAGTTGTATTATA 83676
QY 300 ----- 300
Db 83675 AAGCCAAAGTTTCAGGCTCAGATTTCCCTGCTCCCTGGTGTGTGACAGTGAACACTGCGTG 83616
QY 300 ----- 300
Db 83615 TATTACAAGTCTATGTTTAGCACTTTGCGTATGAGATTGTGTCACTAGTCAGTCACTG 83556
QY 300 ----- 300
Db 83555 GGCTCCAGTGGAACTCCTCGGTGGCTCTCACATCCGTAGGAAATAACCCACAGGT 83496
QY 300 ----- 300
Db 83495 GCAGCAAGCAGGGTCTCAGCCTCAGACACCGTCCAGCTCATCTACTCATGCTTTCAA 83436
QY 300 ----- 300
Db 83435 TTTAGACCTTACTCATTTTGAATGATATAGCTCACCTCAGATCTGCTGACTAGAGAA 83376
QY 300 ----- 300
Db 83375 TTAAGCTGAAGTACGCTGATTACTGGCTTGTGTTTAACTGCTGATTGTTGGTCTT 83316
QY 300 ----- 300
Db 83315 GTAGAATATCAGACAACATAACTCTGAGACAACATAACTCTGACCTTCCCGTACGAATGT 83256
QY 300 ----- 300
Db 83255 TCTACTGCCCTGATTTAGAAATCCATCTGCACATTCCTCTGTGTTCAATGGCCTGATCT 83196
QY 300 ----- 300
Db 83195 GGTAAGCATCAAGTTTCATTTGTTGAAATAGATTGAGCATAGTCTGCTGGTCAGATG 83136
QY 300 ----- 300
Db 83135 TGGAAAGCTGGACGCTTTATTATTATAGCATATGAGAAATGAACGGGCTGTTTACGACTGA 83076
QY 300 ----- 300
Db 83075 GCTCAGAGCTTTGGGAATTTACTGCTATTTATTTTGTGTTGTTTCCAAAAGTATGG 83016
QY 300 ----- 300
Db 83015 AGCCTTTTATGTCATATATAGCCAGNATGATTAAAGTATAGAGATGATTGGCTTCAG 82956
QY 300 ----- 300
Db 82955 ATGTTTGTGATGTGTTGTTCTTCATCATAGTCTTTAAATGACTGGGATCTTTTCCCCAG 82896
QY 300 ----- 300
Db 82895 TATTAGTGTAGTAGAGACCAGTCTCAGGGGGCCAGTTTAAGAGAGAGACGTCCTGCTGA 82836
QY 300 ----- 300
Db 82835 AACAAATTGACGTCGGGACATTTATTGGAATGGAATTAAGCCAGACCTTTGTGCTATA 82776
QY 301 ----- ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGI 319
Db 82775 TTGCAGATGTTCTGTGAGGGCAAGAGCTGGAGAAGAGCCCTGATATTCAGATCATCAA 82716
QY 319 nAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTy 339
Db 82715 CGACGGGAGCTCCACTCTCTCATCATCTCAGAAGCGTTTGAAGAAGACACCGGGCGCTA 82656

QY 339 rSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGI 359
Db 82655 CTCCTGCTTTGCCTCTAACATCTACGCACCGACTCCACGCTCAGCTGAGATCTACGTGA 82596
QY 359 u----- 359
Db 82595 AGTGGTGTCCGTCCTTCCTCAGGCCACAGCTGTCTTTACCTGTGTCTGATGAAGA 82536
QY 360 ----- GlyValSerSerAspSerGlu----- 367
Db 82535 ATGTGCTTTTATTGTGCAGGTGCCTCCTCTGACTCCGAGGGAGAGCATCATTTTGA 82476
QY 367 ----- 367
Db 82475 ACACACAGCCAGTGTGTTTAAAGCCCATTTTCATTTAACTATTATGAGCCACTTTATT 82416
QY 367 ----- 367
Db 82415 TGATCGCAGTATGCTCTTTCTGCTATTTGAAGTGTTCAGCCAAAATCTTTACCCC 82356
QY 368 ----- Gly-AspProAsnLysGluGluMetAsnA 377
Db 82355 CATTTCAAGTCAGGCTCTCTCTGCAGAGCAGGAAGACCTCTGTATCTCTCTGAAGCAG 82296
QY 377 rgIleGlnLysProAsnGluValSerSerProProThrThrSer----- 391
Db 82295 CCACTGAGGAACACAGCAGACGCCGCCCCACCACTAACCAAAATACCCCCAGAAC 82236
QY 391 ----- 391
Db 82235 CTCCTGCAAAATACTTCCAGTCTCAGATTTCTCCAAACCCCGGGAAGCCTGCAACGC 82176
QY 391 ----- 391
Db 82175 AGGTGACCGAGCAGCGGATCCAAAGCTATCTGTGAGGGGATTTCTGCTTTGTCAACAA 82116
QY 392 ----- AlaValIleProPro----- 396
Db 82115 GTGACCCAGAACATCCTGAAAGTCAGACGCCATTTCCACCGTCTCCTTGTCTGCGCCAG 82056
QY 397 ----- AlaValPro----- 399
Db 82055 CGCAGTCCGCTCAGCAGCCACCTCGCTGCTTTTGACGCGAGTTCCATCCCTCCACCTG 81996
QY 400 ----- GlnAlaGlnHisLeuValAlaGlnPro----- 408
Db 81995 CTTTGACACCTCCACCAGGCATCCCACTGAGGTACACACCACCATCAACTTCCGCCCC 81936
QY 409 ----- ArgValAlaThrIleGlnGlnCysGlnS 418
Db 81935 ATATTATTAGATTAGCTCTGTTGTTTAAACCACCTTATCAACTAATGTCAGAAATCAGG 81876
QY 418 erProThr----- AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValP 437
Db 81875 CCCCACCTCTAATATTATGCAAGTCTGAATGGCAGCCTGTCATGGCAGCACCCATAT 81816
QY 437 heThrLysMet----- 440
Db 81815 TCACAAAGGTACGTTCTCTGCGTGTCTCCACTCTACTGTTCTATCTGCTTCCATCTCA 81756
QY 441 ----- LeuGlnAsnLeuSerAlaSerGluGlyGlnL 451
Db 81755 CCATCGACTGTTGTTGTCCTCCAGAACCTTCGAAGAGCTCTATGATCCGAAGGCCAAT 81696
QY 451 euValValPheGluCysArgValLysGlyAlaProSerProLysValGluTrpTyrArgG 471
Db 81695 TGTGGTGTAGATGCGGTGTGAAGAGTACCGTCCCGCGAGTGGCGCTGGTATAGGA 81636
QY 471 luGlyThrLeuIleGluAspSerProAspPheArgIleLeuGlnLys----- 486
Db 81635 ATGAAACATGATTGAGGATTTCCACAGATTTTCAGGATCTTACAGAAAAGTAAGGCATAAT 81576
QY 486 ----- 486

Db	81575	TATGATTCTTGTGGTTGGTTGGTTAGTAGCAGAAAGTAGCTGTGCACGTAGTAG	81516
Qy	486	-----	486
Db	81515	TGGTCATTCAAGTTCAAAAATATTTAATGATCCCTCAAGCACATTATTCTACACACTGGA	81456
Qy	486	-----	486
Db	81455	AACATTAAAAGTGTTTTGAACAAACTGTGCCTGTGTAGAAGCCTGCAGGTATATTGT	81396
Qy	486	-----	486
Db	81395	ATGTGTTTTTGGAACTGCAGTGGAAAGACTTTTTCCAAGGAAATCTTATCTTCTCTTT	81336
Qy	486	-----	486
Db	81335	TAAATGTTTTTTTTCCGTCGTTTTCAC TTGATAAGTATGGAGATGTGTTAAACTGCC	81276
Qy	486	-----	486
Db	81275	TTTGGCATCCCACTCAAATGTTTTTGGCTCTGTGGAAATTAAGATCACAACTGACATC	81216
Qy	487	-----Lys- - - - - :::	487
Db	81215	CTAAGAAACAGTGACTCAGCAGTGATGGGAAC TTTTGGTGT TACTTCTGTTC CAGAG	81156
Qy	488	ProArgSerMetAlaGluPro- 	494
Db	81155	CCCAGATCTCCAGCTGAACCAGGTAAAGCGTTTGAATCCCCACCTTCATGCACAAGGATG	81096
Qy	494	-----	494
Db	81095	TGTGCACCAAGCAGTTGGTTGTGTTAATGATGCTGCAGCCACAGTACATATTGCCTT	81036
Qy	494	-----	494
Db	81035	TTACTGAAGTGTTTATAGTCTAAGCCGAGCCTTGTGCTCTCTTGTATATAAGCCAAG	80976
Qy	494	-----	494
Db	80975	GCTGGTTCCCGTGACTCTAGACCATATAAGGCTTCCTCTCATCTGTATCAAACTGGCC	80916
Qy	494	-----	494
Db	80915	CGGTAGATCACTGTAACCTGATTTAGCTGCAGATAGACTCAAATGCAGAAGAGTTTTTA	80856
Qy	494	-----	494
Db	80855	CCCACAAGTTGTGTCTGTCTCTCACTAATGCATAGTGGAACTCTCTCTAGATAAGAACT	80796
Qy	495	-----GluGlu 	496
Db	80795	TGGTTTTATGTAACAAACAGCATTGTCTCATTTTAAAAACCAATATATATTTTAGAGGA	80736
Qy	497	IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCystThr	516
Db	80735	ATATCACTCTGGTGATTGCTGAAGTTTTTCTGAAGATTCTGGCATCTTTACTTGTTACT	80676
Qy	517	AlaSerAsnLysTyrglyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGlu	536
Db	80675	GCAAGTAACAGTTTGGAACTGTCTCCAGCACAGCTGACATAAGGGTCCAAGGTAATGTT	80616
Qy	537	AspLeu- 	538
Db	80615	TT-CTAAAATTCGTGGCGTCGATTGTGTTATTAAATTCATCTAAPTAACATATTTCT	80557
Qy	539	-----SerAsnAsnGlySerLeuHis ::: :::	545
Db	80556	GTGATTTTCATCAGGGAATGGCAACAGCAACCATGTCAGGCCCTTTGCAGAGTGTACAGGTG	80497
Qy	546	-----SerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnProSer ::: ::: ::: :::::	561

D	B	80496	GAGCCTGGCCAAAGCCAAGAACCCTTCCCCTACGCTTCATCGGTGAACCAACCCAACAGAG	80433
Q	Y	562	-----ProHisSer 	565
D	B	80436	GTGACTGTGCCAACAAGCATCAAGCCCCACTCCAGCACCATTTGCCTGGACCCGCTCATC	80377
Q	Y	565	-----	565
D	B	80376	TCTGTAACGTACGGCTGGAGCCCTTAACACACCTGCACCTCCGCTTGATCCCAACAAC	80317
Q	Y	566	-----GluproSerValGluInPro- :::	573
D	B	80316	TCCAGCATGTTGGACACAGATCCACCATGACACAGTTCACCCAGTGTGCAAAACCTGCAGT	80257
Q	Y	573	-----	573
D	B	80256	CAGACGACGCTACTGCCTCAGCGCTCAACGTTTTAAATCTGTCTACCTCCACCAGAGC	80197
Q	Y	574	-----ProLysProLys 	577
D	B	80196	CCTGTTGAACCCAGTGGAGCTAATCCAGAGCCCCAAAAACACAGGGGGGCTTCTTAC	80137
Q	Y	577	-----	577
D	B	80136	ACCAAACCTTCAACCTTTGTGCTGCTTATGATGCTCCAGTTGGCAGGTTTCCAGATCT	80077
Q	Y	577	-----	577
D	B	80076	TCAGTGTCTTTCTGTGATACCAAAAGTTAAGGATGCCCGGACCATGACAATTGGNAATCAG	80017
Q	Y	578	-----LeugluyValLeuVal 	583
D	B	80016	GTTGTAGTGCTTCTCCGATCCCCCTCCCAACTCTCTGCTTAAAGTCAGGCACCTGACA	79957
Q	Y	584	AnnHisAnsgluProArgSerSerArgIleGlyLeuArgValHisPheasnLeuPro	603
D	B	79956	AACCAAGGACCCCGCTCCAGCTCCAGTCGAGCGCGCTGCGCTGCACTTCAAATTGCGG	79897
Q	Y	604	GluAspAspLysGlyserGluAlaSerSerGluAlaGlyValValThrArgglInThr	623
D	B	79896	GAGGATGAAGAA-----GAACCGCAAAAGTGAC-----ACATCTGTGCTGCT	79855
Q	Y	624	ArgProaspSer**GlnGluArgPheAsnglyGlnAlaThrLysThrProGluProSer	643
D	B	79854	GAA--GATGCTGCTCAG-----	79840
Q	Y	644	PheProvalLysGluProProProValLeuAlalysProLysLeu-----	658
D	B	79839	ATGCTGCCAAGGAACCCACCCTCGAGCTTAAGCCCAACTGTGATGATAACAGATA	79780
Q	Y	658	-----	658
D	B	79779	ATAAAAACAACAGTGGNACCCGATGGCTGCAGTGTGGGAGGAAGAGATTTTT	79720
Q	Y	658	-----	658
D	B	79719	ATAGATGAGTTGTGAACACTGCTCTTGTTACATTAAGGGTGTAAAGAGGTGGGAG	79660
Q	Y	658	-----	658
D	B	79659	GTTCTCTCCACTGCATTTACAAGCAGCTGCAACCGCTACAATTCATAAACGCATCCA	79600
Q	Y	658	-----	658
D	B	79599	CTCAGACTTCAATGTGGCTTTGCGAGCTTCTTAAACAAGGATGGTATTTTCTACTTG	79540
Q	Y	658	-----	658
D	B	79539	TAAGTCTAGCTGTTTGTATGCTCTGATAGTCTTAGATAATGTTGGAAATGAATCTTGC	79480
Q	Y	658	-----	658
D	B	79479	ATGTTTGATTGTCTTATTTTGTCTCATTTGCCITAMAGACAAAATACAATTTAGATAG	79420

Qy 658 ----- 658
Db 79419 ATTAGTCTGAGACATTTTACATGAATTTGATAAATAAATAGTAGTCATTCAAATTTGAAA 79360
Qy 658 ----- 658
Db 79359 GCTGTAGTGTAGTAATAGCAGGGTTCTTTGTTGTATCCATAAAAGGCGCAGCATATTCTT 79300
Qy 658 ----- 658
Db 79299 CACTCTGGGGTTTACATGGATGGTTTATTAGAGGAGTGCAGAAAGCTTCCCCCCTCAC 79240
Qy 658 ----- 658
Db 79239 CTTTGACCATGTATAGTGTGATTTTAAAGCTGGGCCATTTGTTGTTGTTTAAATTGAC 79180
Qy 659 --- AspSerThrGlnLeuGlnGlnLeuHisnGlnValLeuLeuGluGlnHisnLeu 677
Db 79179 AGGGACCCAGCTCAGCTCCAGCTTCTGCACAACCCAGGTCTCTTTGGAGCAGCAGCAGGAG 79120
Qy 678 GlnAsnPro----- 680
Db 79119 ACGNACCACAGACACAACTCAACTCAACTCAATTCACCCACCACCGCAGCTCCAG 79060
Qy 680 ----- 680
Db 79059 GTCCAGCCCGAGGTCAGACCTCCGACAGGTCTATCTTGCTCTCTCAAAATAAACCGG 79000
Qy 681 ----- ProProSerSer----- ProLysGlu 687
Db 78999 GGTCCTCAATCTCCACATTTCCACACCCGACACCTCTCTCTACAGCTCTCCCGACCTCT 78940
Qy 688 PhePro----- Phe***MetThrValLeuAsnSerAsnAla 699
Db 78939 GCCCCTCTCGCTCCNATGGTCAACACCACCTTTCTCCCCCTCTCCAAACCTCTCCTGCT 78880
Qy 700 ProProAlaVal----- 703
Db 78879 CTTCTCAAGCTTGTCATATCTCTCTCTCAGCTAAAAAAGCTCCTACTGGTGACATCC 78820
Qy 703 ----- 703
Db 78819 CTTCCACCTCCCCACAGATAGGAGAGCTGCTCTCTCAGAGCATCCCCCCCCGTAAGTCAG 78760
Qy 703 ----- 703
Db 78759 ATGGATACCTCCATCTGCCCACCTCTGCTGTCTGTCCTCCCGTTGCNAGAGCTGGACCCACA 78700
Qy 704 ----- ThrThr 705
Db 78699 GCAAACTCTCTGACTCCCTGGSCACCAAGCTGAGTGCAGCGCCGTCGATCCACCCCA 78640
Qy 706 Ser***LysGlnValLysAlaProSerSerGln----- 716
Db 78639 TCTCGCAGCAGATACGTCTCCAGCACCAGCCAGCTGAGGACACCCATGTCCTTCTCATG 78580
Qy 717 ----- ThrPheSerLeuAlaArgProLysTyrPhePheProSer 729
Db 78579 AATCTCACACCTCTCTCCCTCCCTTCAACTATGCCCCGTCCAAAAGAAATTCATA----- 78526
Qy 730 ThrAsnThrThrAlaAlaThrValAlaProSerSerProValPheThrLeuSerSer 749
Db 78525 ----- GCTGCACACACTTTCTCCCCAGTCAGGAGCCCCCTCTTAACAGAAATCCCCA 78475
Qy 750 ThrProGlnThrIleGlnArgThrValSerLysGlnSerLeuLeuValSerHisProSer 769
Db 78474 GTACCTCTGCTGCGAGGAATCTGCTGCAGAGCTCAACTCTCTCAGAGGCAAGTTCTCCAC 78415
Qy 770 ValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProPro----- 787
Db 78414 ATTCCACCTTTCTCCCA-----CCCCCGAGGGTT 78385

Qy 788 ----- GlyProThrGluPro----- ThrPro 794
Db 78384 TTTCCACACGGTTCTGATGCTCTCCACACGCTCCATCCTCCTGTGTCTTCCTCC 78325
Qy 795 ProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerPro 814
Db 78324 ACACCGGATCGTCTCTCTCTCTGAGC---AGCTCGTTTCGCCATCCGAGCTCAGTCGCGCC 78268
Qy 815 IleProVal---SerProThrSerArg----- IleGlnAsnProVal 827
Db 78267 CTACATGCATCTCTCCACACCTCCAGAGCTCCACTCGAGCCCTATTTCAGATCCCGTG 78208
Qy 828 AlaPheLeuSerSerValLeuProSerLeuProAlaIleProProThrAsnAlaMet*** 847
Db 78207 GCATTTCTAAGTTCGCTCTCCATCCCTCTCCCTGGCTCAGCCCAACCACTCGATGGC 78148
Qy 848 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAla----- 861
Db 78147 CTGCCGAAGGGGCTCCGTC-----GGGTACGCTTTTCCCAATGACCTGTCA 78100
Qy 862 ----- LysLys 863
Db 78099 TTCGTGACCGATGGTGCACGCCGATTCACCTCTTTCATGCTCTTCAGGCTTAAAAAG 78040
Qy 864 AsnThrLysSerProGlnProValAsnAspAsnIleArgGluThrLys----- 880
Db 78039 AGCTCCCAAAGGCAGCAGCTCAATCAACAGATAATGTCTGTAAGCAAGAGAGATTCTC 77980
Qy 880 ----- 880
Db 77979 CATGAATTTGAAAAAGCTTCGATTTAGAGAGGAAACTCCACATTTTGCACATCAACAG 77920
Qy 880 ----- 880
Db 77919 GTACTATATCTCTCATGATTTATTAGCTGATTTCCCTCCATCGGTAGCAGTGGTAAAT 77860
Qy 880 ----- 880
Db 77859 GGAATGATTTCAAAGACAAAAAATAATACCTAAGACAAAAAATAATAGCAGTGAAGTC 77800
Qy 880 ----- 880
Db 77799 CTCCTTTATCGAAATGCTGTCTGATGATTTTCATTTTCAATTACATCCCTGTTTATAT 77740
Qy 880 ----- 880
Db 77739 TATGTCTAATGTCTGGGCATTTGGATGTTGGTAGATTTGAATTTGACACAAGGTGAAG 77680
Qy 880 ----- 880
Db 77679 TATGTGAATGCTAGCTGCCATCTAACCAATGCTGGTATCTGTTTAACTCTTTTAAAGG 77620
Qy 880 ----- 880
Db 77619 ATAATTTGGTTGATCTTAATAATGTTATATTAAGGAAGCTGCCCTCAACAATGTCTGA 77560
Qy 880 ----- 880
Db 77559 GAAATGAGGAAGAACCCTGTTCAAGCCGACGTTCTGACTTCATCTTGTACTGTCAAT 77500
Qy 880 ----- 880
Db 77499 TACCGTGTGTACCATCACCTAACGTATGCTAGTAGAGCCCAACTTGTGTTCTTCCA 77440
Qy 880 ----- 880
Db 77439 TCAGAATGTACAATGTGTGTATGTAGTGTTCATCCATCCCTGCTCTCATAGAAG 77380
Qy 880 ----- 880
Db 77379 CTGAGTGTGGGGGAAAAACAGCAGCAGACCTGTGCGATCAAACTTTCCTGCTACTGTC 77320
Qy 880 ----- 880

Db 77319 GTTAGCTATGACGAGGTACAAGGATCCTCCACAGCGGTGCTCGCGCTGTACGGT 77260
Qy 881 -----Asn-AlaValIleArgAsp----- 886
||| |||||::: :::
Db 77259 CATTAATAATGCACTAAAGAGAACCGCAGTTCTCTAAACTTATCTATAAAGCT 77200
Qy 886 ----- 886
Db 77199 TTATCTAATATTACACATAACATTTACTAATAATTACAGGGAATTCAACTGATTTTAGGGTA 77140
Qy 886 ----- 886
Db 77139 AGAATCCATTAAACTTATTGAATGTTGCTGCTTCTATACAATAAATCACTCATCGAAACGTC 77080
Qy 886 ----- 886
Db 77079 TTGTTTCCCTTTTGTGATTTTTTTTTCACATAGAAAATACCTTATTTCCCAACCACT 77020
Qy 886 ----- 886
Db 77019 GTTATTAAATTTGCTTTTGTCTGCTGCTGAGGTGATGTAAGCATTTCTTAGTGTGACTC 76960
Qy 886 ----- 886
Db 76959 AAACAAGGAAGTGAGTGTGGATGTTTCTTTTAATTTTGAGTGTCTCAGTTTCATCATC 76900
Qy 886 ----- 886
Db 76899 TCTGTTCTCTGTTTCTTCCGCCAGTGTGCCAGGCATGTCTGCTGCATGTCTATGTGAC 76840
Qy 886 ----- 886
Db 76839 ATGCCAGTTTGACAGTTCGCTTATAAATTCCTACTCTAATGCTACAGTTACTAGATGAC 76780
Qy 886 ----- 886
Db 76779 TGCACATAATCAGACTTGTAGTAAGAGCTTGTGCTGCATCTTTGATTACGTTTGTATGGAC 76720
Qy 886 ----- 886
Db 76719 AGTTGCCAAAATCCAAAACACATCTCCTTATTCCTGTTATTCCTGTTATTAATCTATCTAGACAC 76660
Qy 886 ----- 886
Db 76659 TTCTGATATAATTTGGAGATCTTTGGAGAGCATAAAGTGGNAAGATTGTAAAAATTCCT 76600
Qy 886 ----- 886
Db 76599 ATAACAGAAGAAGAGGCTTTGGTGAAGTAAATAAATAATATATTTGGGCAACCAACCA 76540
Qy 886 ----- 886
Db 76539 TAAGAGTTGCTATCATCCATCTGCAAAATCCTGTTGTAAAGTCITTTATATAAACTTTATCA 76480
Qy 886 ----- 886
Db 76479 GTTTTTCAATGTTGCCAAGCTAATTTGGACCACAGACAGACAATCATCTAACACCATCA 76420
Qy 887 -----LeuGlyLysIle 891
||| |||||::: |||||
Db 76419 CCAGCTTAGGAGTTATAGGAGGCTTTAGGAGGTACCCGCAGACTAACTAGGACGTAATAAT 76360
Qy 891 eThrPhe-----SerAspValArgProAs 899
||| |||||::: :::
Db 76359 CACCTTTAAAGCTAACTGACAGTTAGACGGTGTGATACAGGGCCACAGAGGAACACAGA 76300
Qy 899 n----- 899
Db 76299 GGACTGAACAGAGGAGTGACTTCAATTTTAAAGAGAAAGCAAAATATATGCGCTATACCT 76240
Qy 899 ----- 899

Db 76239 TAAATGATCAAAATCCCAACAGATTAATCCTGTTGGAAACAGGGAAGATGCAATTTGAT 76180
Qy 899 ----- 899
Db 76179 TTTAACTCTGAGCATCAGCTTCTAAGATACGGTATGATGTTACTGTCCAGCACAGTTCA 76120
Qy 900 -----GlnGlnG 902
||| |||||
Db 76119 GTGCTGGATGCATCATTTGGCTTTGACTTTACATCATGGGTTCTTTTTCCTCCAGGA 76060
Qy 902 uThrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluAr 922
||| |||||::: |||||
Db 76059 GTCAAAAGTGTCCAGTTTGTGACAGAGGCTCATGAGTGAATAGATTTCTGTTAGAAGC 76000
Qy 922 gThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGlyLysCy 942
||| |||||::: |||||
Db 75999 AACCCCTGTGGAGGAATCGGATGATGAAGTCAACATGATGATGCCCACTGGGAATG 75940
Qy 942 sIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVa 962
||| |||||::: |||||
Db 75939 TATAGCCCATCTTTTGAAGAAGCTGAAGAAGCTTCAAGCTACGGAGGCGTCCCTGT 75880
Qy 962 lThrPheThrCysLysIleValGlyIleProValProLysVal----- 976
||| |||||::: |||||
Db 75879 CACCTTCTCATGTAATAATATTGGAATCCTGTTCCAAAGGT-GAGGGAGACACAACGT 75821
Qy 977 -----TyrTTPheLy 980
||| |||||
Db 75820 AACCCCCCTTAGGAACAATACATTTATCTGTAAATCCTCAACAGGTTTACTGGTTCAA 75761
Qy 980 sAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluGlyAspG 1000
||| |||||::: |||||
Db 75760 AGATGGCAAGCAGATCTTAAAGAAGACGCTCCATACAGGAAGATAAGAGGGGGTGG 75701
Qy 1000 yThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMe 1020
||| |||||::: |||||
Db 75700 AACCTGTCTTTACATATTGAGTGCACGACCAGTGTATGATGGAACACTACACGTAAT 75641
Qy 1020 tAlaAlaAsnProGln----- 1025
||| |||||
Db 75640 GCGCGCTAATCCACA-GGTAGTACACTTCTGTCTGAGTTAGCATGAGCTTCCCTCTGAA 75582
Qy 1026 -----GlyArgIleSerCysSerGlyHisLe 1034
||| |||||::: |||||
Db 75581 AGCAAAATCCAAACTCTTGTTAATCACCATCAGGAGCGGATCAGCTCAGGTCAATT 75522
Qy 1034 uMetValGlnSerLeuProIleArgSerArgLeuThr-----Ser-AlaGlyGlns 1051
||| |||||::: |||||
Db 75521 GATAGTCCAAACGGGACCTCTCAGAACCCGACTGTCCCATTTTTCATTTCTCAGAGGTGAGC 75462
Qy 1051 er----- 1051
Db 75461 AGAAGTCATTTGCAAGTGGAGAACTGGATCATATTTGACATCATATATCATGTCCTCTCT 75402
Qy 1052 -----HisArgGlyArgSerArgValGlnGluArgAspLysGluProLeuG 1067
||| |||||::: |||||
Db 75401 GCTTACACATCTTTTTCAGGGGGCAATCCTCGTACGGGATGCTGAGGGTGAGCAAAACC 75342
Qy 1067 lnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluG 1087
||| |||||::: |||||
Db 75341 AGGAGCGCTTTTTCAGCTCATTTTCTTCAAGCTCCAGGAGACATGCTGGCTCATGAGG 75282
Qy 1087 lyArgLeuCysArgLeuAspCysLys----- 1095
||| |||||::: |||||
Db 75281 GAAAACTGTGACGCTAGACTGTAAAGTGAAGATCAGGCCAGGGGACCACAGAAATATG 75222
Qy 1096 -----v 1096
Db 75221 GGAATGGTTTTCATATTATGTTTTCCTCTCATTTGATTTCACTGTTTCTTCTCTCCAGG 75162
Qy 1096 alSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProA 1116
||| |||||::: |||||
Db 75161 TGAGTGGTCTACCCATCCACAGTTGATGTGTTGGTCAACGGGAACCCATCTACTACAG 75102

Db 52 GGTCCGGAGCCGCTGCAGCTCCCG-----CTCGTCCGGACGCGGAATCGGGC 102
QY 655 -----LysProLysLeuAspSerThrGlnLeuGlnLeuHis 667
Db 103 AGCAGCGGGAGCGCGCCGAGAGCCGAGG---GACCCTCTGAAGCTCCAGCACTCCAG 159
QY 668 AsnGlnValLeuLeuGlnHisGln---LeuGlnAsnProProPro-----Ser 683
Db 160 AACCAATCCGACTGGAGCAGGAGCGCGCTCGGCAGCCTCCGCCAGCCCGCGCAGC 219
QY 684 SerProLysGluPheProPhe----- 690
Db 220 GCAGCGCCCTCGCCGCCCTTCCCGCGCGCCGCGCTTCCCGAGCTCGCGGCTGCAGC 279
QY 691 -----***MetThrValLeuAsnSerAsnAlaProProAlaValThr 704
Db 280 CCGCGCGCTCCCGGAGCCCATGAGCGCGCTGCCCTCCCGCTCCGCGCCCGCATG--- 336
QY 705 ThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeuAlaArgProLys 724
Db 337 -----CAGTCTCTCCGCTCCTTCAACTACGCGCGCCCAAG 372
QY 725 TyrPhePheProSerThrAsnThrThrAlaAlaThr-----ValAlaProSerSerSer 742
Db 373 CAGTTCATCCCGCGCAGAACCTCGGCGCGCTCGGCGCCACGCGCGCGCTCCAGC 432
QY 743 ProValPheThrLeuSerSerThrProGlnThrIleGln-ArgThrValSerLysGluSe 762
Db 433 CCC-----AGCTCGTCCAGCTCCCTCGCCCATGTCGCCGAGCGGAGGAGTTCGGC 486
QY 762 rLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAs 782
Db 487 CGCGCCCCGTGCGCGCTTCG-----CGCAGCGCTTCGCGCTGAGCGCGAGGCC 537
QY 782 nGluProLeuProGlyProThrGluProThrProProPhePheThrPheSerIlePr 802
Db 538 CCGTGGGCTCTCTCTCG-CCGTGCGCGCGCGCGCGCGCGCGCGCTCTTCAGC----- 591
QY 802 oSerGlyAsnGlnPheGlnProArgCysValSerProIlePro----- 816
Db 592 -----CCAGGCTGCTTCCCGGTCGCGAGCTGTCCTCCACGTGCC 632
QY 817 -----ValSerProThrSe 821
Db 633 GCGGCCACCGCGCTCCCGAGCGCGGACAGCGCTCCCACTGCTCGTCCGCTGCCAC 692
QY 821 rArgIle-----GlnAsnProValAlaPheLeuSerSerValLeuProSerLeuPr 838
Db 693 CCGCTTCGCGCAGCGCGCGCGCGCGCTTCTCAGCGCTGCTGCGCTCGCAGCC 752
QY 838 oAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerG1 858
Db 753 GCGCGCGCGCGCTCAACCGCTTGGGGCTGCCCCAAG-----GGTGTACCCCCCGC 803
QY 858 nGlyLeuAlaLysAsnThrLysSerProGlnProValAsnAspAsnIleArgG1 878
Db 804 AGGATTCCAAAGAGGCCAGTAGAAGTCTAGAAATAGCTCCGATGAGGAAATTCAGG 863
QY 878 uThrLysAsnAlaValIleArgPheLeuGlyLysLysLysLysLysLysLysLysLys 898
Db 864 CACAAGGATGCTGTTATTCAAGACTTGAACGAAACTTCGCTTCAAGGAGGACCTCTCT 923
QY 898 oAsn----- 899
Db 924 GAACAATGGCCGCGAGGTTAAACATACGAAGAAGAAATGGCTGCTGCAGTGTAGTGC 983
QY 900 -----GlnGlnGluTy 903
Db 984 TGACAGTGAACCTGCTCTTTAATATTCAGGAGCCAGAGAGGAAACACTTAATCAGGA 1043
QY 903 rLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgTh 923
Db 1044 CAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAATAGGTACAGGCTAGAAAGG 1103

QY 923 rProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysI1 943
Db 1104 TCTGTGGATGAATCAGTGAATGAATTCAGTATGGAGATGTGCTGTGGAATAATGAAT 1163
QY 943 eAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValTh 963
Db 1164 GGCACCATTTTGTGAGATGAAGCTGAACATTAACAAGATCTTTGAGGAATGCCAGTAAC 1223
QY 963 rPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysaspGlyLy 983
Db 1224 TTTTCATGTAGATGGCTGGAATCCAAAGCCAAAGATCTATTGGTTTAAAGATGGAA 1283
QY 983 sGlnIleSerLysArgAsnGluHisCysLysMetArgGluGluGluGluGluGluGlu 1003
Db 1284 GCAGATCTCCAAAGAGTATCACTACACCATTCAAAGAGATCTCGATGGACCTGCTC 1343
QY 1003 rLeuHisIleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMetAlaLaas 1023
Db 1344 CTTCCATACACAGCTCCACCTAGATGATGATGGAATATATAAATTTATGGTGCATA 1403
QY 1023 nProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArgSe 1043
Db 1404 CCCTCAGGGCGCATCAGTTGTACTGGACGGCTAATGGTACAGGCTGTCAACCAAGAGG 1463
QY 1043 rArgLeuThrSerAlaGlyGlnSerHis-----ArgGlyArgSerAr 1057
Db 1464 TCGAAGTCCCGGCTCTCCCTCAGGCCATCTCATGTGAGAGGCTCGTCTAGATCAAG 1523
QY 1057 gValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuG1 1077
Db 1524 GGACAGTGGAGAGAAAAAATGAACCAATTCAGGAGCGATTCTTCAGACCTCCTCTTTC 1583
QY 1077 nAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSe 1097
Db 1584 GGCTCTCTGGAGATCTGACTGTTCAAGAAGAAACTCTGCAGAAATGAGCTGCAAGTCA 1643
QY 1097 rGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAl 1117
Db 1644 TGGTTACCAACCCAGATCTAAGCTGGCACTAGTAGGAAGAGCCCTAGCCCTGTACAG 1703
QY 1117 aSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuTh 1137
Db 1704 TGCTCAACAAGATGCTGCTGCTGAGAACGGGTGCTCACTCTCTGATCATAGAGCCAGTCA 1763
QY 1137 rGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPh 1157
Db 1764 GTCAGTGTATGCCGCTATACACATGTATAGCTACCAAGCGAGCAGCAGACACTCAT 1823
QY 1157 eSerLeuGluLeuSerValValAlaLysValLysLysAlaProValIleLeuGluLy 1177
Db 1824 CAGCTGGAGCTGTGTTGCTGCTAAAGAGACACAAACCCCTGTGTTTATTGAGAA 1893
QY 1177 sLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleG1 1197
Db 1884 GCTCCAAAACACAGGAGTGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1943
QY 1197 yMetProProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgG1 1217
Db 1944 AGTGCCACCTCAGATATTTTGAAGAAGAAAGTAACTCACTCACTCACTCACTCACTCA 2003
QY 1217 uArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLy 1237
Db 2004 CCGAGTGCAGTGCAGCAGGACACACCGCTACATCTGCTGCTCATTCAGGAGGCCAC 2063
QY 1237 sLysSerAspAlaGlyTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCy 1257
Db 2064 AAAAGAAGATGCTGGTGGTATCTGTGTACGCCAAGAAATGAACAGGAGTGTGTCTG 2123
QY 1257 sThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis-----GlnIleProProProme 1275
Db 2124 TACTGCCAGGCTGAGCTTTTACACCCAGTGGCATCAGCAGTGCAGCAGCAGCAGCAGC 2183

Qy	1275	tSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePh	1295
Db	2184	AAAAATGAGGCGCCCTCAGCAGCTCGCTATGCAGCACTTTCCGACACAGGACTAGACATCAA	2243
Qy	1295	eSerAlaPheSerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValVa	1315
Db	2244	AGCAGCGTTCCAACT---GAGGCCAACCCATCTCACCTGACACTGAATACTGCCTTGGT	2300
Qy	1315	lGluSerAspGluLeu	1320
Db	2301	AGAAAGTCAGGACCTG	2316
RESULT	11		
LOCUS	AC024258	160658 bp	linear PRI 14-MAY-2002
DEFINITION	Homo sapiens chromosome 10 clone RP11-297N15, complete sequence.		
ACCESSION	AC024258		
VERSION	AC024258.9	GI:20564414	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Smith,D.R.		
JOURNAL	Genome Therapeutics Corporation Sequencing Center: Human Genome		
REFERENCE	Sequence Data		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 160658)		
JOURNAL	Smith,D.R.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (28-FEB-2000) Genome Therapeutics Corporation, 100 Beaver		
TITLE	Street, Walcham, MA 02453, USA		
JOURNAL	3 (bases 1 to 160658)		
REFERENCE	Smith,D.R.		
AUTHORS	Direct Submission		
TITLE	Submitted (13-NOV-2001) Genome Therapeutics Corporation, 100 Beaver		
JOURNAL	Street, Walcham, MA 02453, USA		
REFERENCE	4 (bases 1 to 160658)		
AUTHORS	Smith,D.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-MAY-2002) Genome Therapeutics Corporation, 100 Beaver		
COMMENT	Street, Walcham, MA 02453, USA		
FEATURES	On May 14, 2002 this sequence version replaced gi:16905139.		
source	Location/Qualifiers		
	1..160658		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/clone="RP11-297N15"		
	/clone_lib="RPC1-11"		
BASE COUNT	50744 a 31675 c 30448 g 47791 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	6.9e-49	Length:	160658
Score:	1553.00	Matches:	302
Percent Similarity:	99.34%	Conservative:	0
Best Local Similarity:	99.34%	Mismatches:	2
Query Match:	22.54%	Indels:	0
DB:	9	Gaps:	0
US-09-818-990B-2 (1-1320) x AC024258 (1-160658)			
Qy	1	MetGlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr	20
Db	110106	ATGCAAGACGACAGCAGAVAGAAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT	110165
Qy	21	LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn	40
Db	110166	TTAGCTGAACACGACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCCTCCAAC	110225
Qy	41	ProCysHisPheGlySerProSerGlyAlaAlaGluGluGlyGlvGlnAspAspLeu	60

AUTHORS
TITLE
JOURNAL
COMMENT

Smith, D.R.
Direct Submission
Submitted (22-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Aug 24, 2000 this sequence version replaced gi:8439815.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg198
----- Summary Statistics
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 153700 bases at least Q40
Consensus quality: 164398 bases at least Q30
Consensus quality: 167130 bases at least Q20
Insert size: 173464; sum-of-contigs
Quality coverage: 4.1x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.

1
* 1146 1245: contig of 1145 bp in length
* 1246 1245: gap of unknown length
* 2414 1246: contig of 1169 bp in length
* 2415 2514: gap of unknown length
* 2515 3864: contig of 1350 bp in length
* 3865 3964: gap of unknown length
* 3966 5078: contig of 1114 bp in length
* 5079 5178: gap of unknown length
* 5179 6647: contig of 1469 bp in length
* 6648 6747: gap of unknown length
* 6748 7926: contig of 1179 bp in length
* 7927 8026: gap of unknown length
* 8027 9412: contig of 1386 bp in length
* 9413 9512: gap of unknown length
* 9513 10563: contig of 1051 bp in length
* 10564 10663: gap of unknown length
* 10664 11995: contig of 1332 bp in length
* 11996 12095: gap of unknown length
* 12096 13309: contig of 1214 bp in length
* 13310 13409: gap of unknown length
* 13410 15172: contig of 1763 bp in length
* 15173 15272: gap of unknown length
* 15273 17973: contig of 2701 bp in length
* 17974 18073: gap of unknown length
* 18074 21419: contig of 3346 bp in length
* 21420 24881: contig of 3362 bp in length
* 24882 24981: gap of unknown length
* 24982 31876: contig of 6895 bp in length
* 31877 31976: gap of unknown length
* 31977 49015: contig of 17039 bp in length
* 49016 49115: gap of unknown length
* 49116 66286: contig of 17171 bp in length
* 66287 66386: gap of unknown length
* 66387 86201: contig of 19815 bp in length
* 86202 86301: gap of unknown length
* 86302 132480: contig of 46179 bp in length
* 132481 132581: gap of unknown length
* 132581 175265: contig of 42685 bp in length.

FEATURES

Location/Qualifiers
1. .175265
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="10"
/clone_lib="RPC1-11"
1. .1145
/note="assembly_name:Contig1"
misc_feature
1246. .2414
/note="assembly_name:Contig11"
misc_feature
2515. .3864
/note="assembly_name:Contig12"
misc_feature
3965. .5078
/note="assembly_name:Contig15"
misc_feature
5179. .6647
/note="assembly_name:Contig18"
misc_feature
6748. .7926
/note="assembly_name:Contig20"
misc_feature
8027. .9412
/note="assembly_name:Contig23"
misc_feature
9513. .10563
/note="assembly_name:Contig25"
misc_feature
10664. .11995
/note="assembly_name:Contig27"
misc_feature
12096. .13309
/note="assembly_name:Contig28"
misc_feature
13410. .15172
/note="assembly_name:Contig29"
misc_feature
15273. .17973
/note="assembly_name:Contig30"
misc_feature
18074. .21419
/note="assembly_name:Contig31"
misc_feature
21520. .24881
/note="assembly_name:Contig32"
misc_feature
24982. .31876
/note="assembly_name:Contig33"
misc_feature
31977. .49015
/note="assembly_name:Contig34"
misc_feature
49116. .66286
/note="assembly_name:Contig35"
misc_feature
66387. .86201
/note="assembly_name:Contig36"
misc_feature
86302. .132480
/note="assembly_name:Contig37"
misc_feature
132581. .175265
/note="assembly_name:Contig38"
misc_feature
BASE COUNT 53132 a 34451 c 33650 g 52102 t 1930 others
ORIGIN

Alignment Scores:
Pred. No.: 7.59e-49 Length: 175265
Score: 1553.00 Matches: 302
Percent Similarity: 99.34% Conservative: 0
Best Local Similarity: 99.34% Mismatches: 2
Query Match: 22.54% Indels: 0
DB: 2 Gaps: 0
US-09-818-990B-2 (1-1320) x AC063964 (1-175265)

QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 162072 ATGCAGACGACAGCATAGAGCTTCTACTTCCATATCTACGCTTCTAAGAGAGAGCTAT 162131
QY 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 162132 TTAGCTGAACACAGACATCGGGGAACAATGAGAGAGTCCAGCGAGCCCTCTCTCCAAC 162191
QY 41 ProCysHisPheGlySerProSerGlyValaAlaGluGlyGlyGlyGlnAspAspLeu 60
Db 162192 CCTTGCATTTCCGACAGTCTCTCTGCGGCGCTCAAGAGGCGGAGCCCAAGATGACCTT 162251
QY 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 162252 CCAGATCTTTCAGCCTTCTGAGCCAAGAGAATAGACGAAAGTGTCATTTTGGCAAGA 162311

```
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaAArgLysArgLeu 100
|||||
Db 162312 CTGGCCATCAATAGACCCCTTTGGAGAGGACGATGAACATCAAGCTAGAAAACGACTT 162371

Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPhcCysGlnAsp 120
|||||
Db 162372 TCTCCTGATCAGATGAACACACTCACCTAACTTAAGTTTGTAGGCCTAACTTCTGCCAGGAT 162431

Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
|||||
Db 162432 AACCTCGAAGTCCACGACGCTCTTAAGAAAGCCCGGAGGAGGAAAGGCCACACATAT 162491

Qy 141 CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
|||||
Db 162492 TGTCTGAAACCCAGTCCAAAAGATATTTTAAATAGGCTGCCGACTTCATTGAAGAG 162551

Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
|||||
Db 162552 CTATCCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGTGGCTGCAAAAAC 162611

Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
|||||
Db 162612 CACAAGAGTAACCTGGAATCTCAAAACAAAGTTATGCAGGAAACAGCTCCAGTTTCTCA 162671

Qy 201 AspLeuSerGluArgGluArgSerValProIleProIleProAlaAspThrArg 220
|||||
Db 162672 GATCTGTGAGAAACAGAGAAAGATCTTCTGTCTCCATCCCTATCCCTGCGGATACCAGG 162731

Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
|||||
Db 162732 GATAATGAAGTGAATCAGCCCTCGAAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 162791

Qy 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
|||||
Db 162792 GCTCCCAAGTGGGGGGTGGTGGAGACACTACACACGAGGGTCTTCCCTCTCATCTCTGTAC 162851

Qy 261 TyrGluGluProLeuGluGlnProProArgPheThrGlnLysLeuArgGluVal 280
|||||
Db 162852 TATGAGAACCTCTGGGGCAACCTCCCGGTCTACTCAAAAGTTACGAGCAGAGAGTT 162911

Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
|||||
Db 162912 CCAGAAGAACTCAGTACAGTTGGATTGTCATAGTGGTAGGAATTCACCACTCAAGTA 162971

Qy 301 ArgTrpTyrCys 304
|||||
Db 162972 AGGTAAATAATGT 162983
```

```
RESULT 13
AC024460/c
LOCUS AC024460 180707 bp DNA linear HTG 28-MAR-2000
DEFINITION Homo sapiens clone RP11-287P14, WORKING DRAFT SEQUENCE, 22
unordered pieces.
ACCESSION AC024460
VERSION AC024460.2 GI:7331497
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180707)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
```

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 28, 2000 this sequence version replaced gi:7108256.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L7059

----- Summary Statistics
Center clone name: 287_P_14

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 168456 bases at least Q40
Consensus quality: 174769 bases at least Q30
Consensus quality: 176914 bases at least Q20

Insert size: 188000; agarose-fp
Insert size: 178607; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1528: contig of 1528 bp in length
* 1529 1628: gap of 100 bp
* 1629 2914: contig of 1286 bp in length
* 2915 3014: gap of 100 bp
* 3015 5061: contig of 2047 bp in length
* 5062 5161: gap of 100 bp
* 5162 7974: contig of 2813 bp in length
* 7975 8074: gap of 100 bp
* 8075 11732: contig of 3658 bp in length
* 11733 11832: gap of 100 bp
* 11833 14787: contig of 2955 bp in length
* 14788 14887: gap of 100 bp
* 14888 19746: contig of 4859 bp in length
* 19747 19846: gap of 100 bp
* 19847 23870: contig of 4024 bp in length
* 23871 23970: gap of 100 bp
* 23971 28748: contig of 4778 bp in length
* 28749 28848: gap of 100 bp
* 28849 35292: contig of 6444 bp in length
* 35293 35392: gap of 100 bp
* 35393 40357: contig of 4965 bp in length
* 40358 40457: gap of 100 bp
* 40458 45936: contig of 5479 bp in length
* 45937 46036: gap of 100 bp

TITLE
JOURNAL

COMMENT

```

* 46037 52180: contig of 6144 bp in length
* 52181 52280: gap of 100 bp
* 52281 57511: contig of 5231 bp in length
* 57512 57611: gap of 100 bp
* 57612 65019: contig of 7408 bp in length
* 65020 65119: gap of 100 bp
* 65120 73479: contig of 8360 bp in length
* 73480 73579: gap of 100 bp
* 73580 81893: contig of 8314 bp in length
* 81894 81993: gap of 100 bp
* 81994 92063: contig of 10070 bp in length
* 92064 92163: gap of 100 bp
* 92164 103396: contig of 11233 bp in length
* 103397 103496: gap of 100 bp
* 103497 112935: contig of 9439 bp in length
* 112936 113035: gap of 100 bp
* 113036 126895: contig of 13860 bp in length
* 126896 126995: gap of 100 bp
* 126996 180707: contig of 53712 bp in length.

```

FEATURES

source

```

1. .180707
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-287P14"
/clone_lib="RPC1-11 Human Male BAC"

1. .1528
/note="assembly_fragment"
1629. .2914
/note="assembly_fragment"
3015. .5061
/note="assembly_fragment"
5162. .7974
/note="assembly_fragment"
8075. .11732
/note="assembly_fragment"
11833. .14787
/note="assembly_fragment"
14888. .19746
/note="assembly_fragment"
19847. .23870
/note="assembly_fragment"
23971. .28748
/note="assembly_fragment"
28849. .35292
/note="assembly_fragment"
35393. .40357
/note="assembly_fragment"
40458. .45936
/note="assembly_fragment"
46037. .52180
/note="assembly_fragment"
52281. .57511
/note="assembly_fragment"
57612. .65019
/note="assembly_fragment"
65120. .73479
/note="assembly_fragment"
73580. .81893
/note="assembly_fragment"
81994. .92063
/note="assembly_fragment"
92164. .103396
/note="assembly_fragment"
103497. .112935
/note="assembly_fragment"
clone_end:77
vector_side:left"
113036. .126895
/note="assembly_fragment"
126996. .180707
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

```

```

BASE COUNT  55951 a 33545 c 34680 g 54429 t 2102 others
ORIGIN
Alignment Scores:
Pred. No.:      7.85e-49      Length:      180707
Score:          1553.00      Matches:    302
Percent Similarity: 99.34%    Conservatives: 0
Best Local Similarity: 99.34%  Mismatches:  2
Query Match:    22.54%      Indels:     0
DB:             2           Gaps:       0

```

US-09-818-990B-2 (1-1320) x AC024460 (1-180707)

```

QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
|||||
Db 154380 ATCAAGACGACACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 154321
QY 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
|||||
Db 154320 TTAGCTGAACACGACATCGGGAAACAATGAGAGGAGTCGAGCGGAGCGCTCTCCAAC 154261
QY 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu 60
|||||
Db 154260 CCTTGGCATTTCGGCAGTCCTTCTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 154201
QY 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuLeuAspGluSerValAsnLeuAlaArg 80
|||||
Db 154200 CCAGATCTTTCAGCCCTTCTGAGCCCAAGAAGAATTAGACCAAAAGTGCAATTTGGCAAGA 154141
QY 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
|||||
Db 154140 CTGGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT 154081
QY 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
|||||
Db 154080 TCTCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGGCTAACTTCTGCCAGAT 154021
QY 121 AsnProArgSerProThrSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
|||||
Db 154020 AACCTCGAAGTCCCACCACTCTAAAGAAAGCCCAAGAGGCAAAAGCCACAGATAT 153961
QY 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
|||||
Db 153960 TGTTCCTGAACCCAGTCCAAAAAAGTATTTTAAATAAAGCTGCGGACTTTCATTGAAGAG 153901
QY 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
|||||
Db 153900 CTATCTCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCTGCTGCTGCAAAAAC 153841
QY 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
|||||
Db 153840 CACAAGAGTAACTGGNAATCTCAAAACAAGTTATGCGAGGAAACACAGCTCCAGTTTCTCA 153781
QY 201 AspLeuSerGluArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
|||||
Db 153780 GATCTGTCAAAAAGACGAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCAGG 153721
QY 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGlnLysArgArgGluAlaGluGln 240
|||||
Db 153720 GATAATGAAGTGAATCACGCCCTTGGACACAGAGCAAGCGGCTGGAAGCGGAGCAG 153661
QY 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
|||||
Db 153660 GCTGCCAGTGGGGCGGCTGGTGGAGACACACACACAGGGTCTTCCCTCTCATCTCTGTAC 153601
QY 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
|||||
Db 153600 TATGAAGAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGAGCAGAGAGATT 153541
QY 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
|||||
Db 153540 CCAGAGGAACCTCGAGTACAGTTGGATTGCATAGTAGTAGGAATTCACCACTTCCACCACTCAAGTA 153481

```

```

Qy 301 ArgTrpTyrCys 304
Db 153480 AGCTAAAAATGT 153469

RESULT 14
BC013867
LOCUS BC013867 1664 bp mRNA linear PRI 10-SEP-2001
DEFINITION Homo sapiens, similar to palladin, clone MGC:10248 IMAGE:3845400,
          mRNA, complete cds.
ACCESSION BC013867
VERSION BC013867.1 GI:15530190
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 1664)
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Institute for Systems Biology
          http://www.systemsbio.org
          contact: amadan@systemsbiology.org
          Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
          Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 14 Row: m Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4689129.

FEATURES
source
1..1664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:10248 IMAGE:3845400"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
387..1544
/codon_start=1
/product="Similar to palladin"
/protein_id="AAH13867.1"
/db_xref="GI:15530191"
/translaton="MAPFFEMKHKYKIFEGMPVTFCTRVAGNPKPKIYWFKDKQIS
PKSDHYTIQRLDGTCSLHTTASTLDDGNYVTIMANPOGRISCTGRMLVMQAVNQRGR
SPRSPGHVPRPRSRDSQDENEPIDQFRPHFLOAPGDLTVQEGKLCMDCKV
SGLPTDLSQLDGKQVPRPSAHKMLVRENGVHSLIEPVTSRDAGLYTCIATNRAGQ
NFSLELVAAKAHAHPVPVIEKQNTGVADGVPVLECKRVGPPPPQIFWKXENESL
THSTDVSMHQDNHSGYICLLIQGATKEDAGWTVTSARKEAGIVSCVTRALDDVYVYQWQQ
SQSTPKPKRPSRYAALSDDGLDIAKAFQPEANPSHLETLNLTALVESEDL"

BASE COUNT 514 a 397 c 406 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 3,76e-48 Length: 1664
Score: 1476.00 Matches: 278
Percent Similarity: 70.51% Conservative: 83
Best Local Similarity: 54.30% Mismatches: 107
Query Match: 21.42% Indels: 44
DB: 9 Gaps: 4

US-09-818-990B-2 (1-1320) x BC013867 (1-1664)

```

```

Qy 852 AlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProVal 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 GCCAAGGGTGTCAACCCCGCAGGATTTCCAAAGAAAGCCAGTAGACTGTAGTAAGACC 68

Qy 872 AsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIle 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 TCCGATGAGGAAATTCAGGCACAAGGATCTGTTATTCAAGACCTTGAACGAAACTTT 128

Qy 892 ThrPheSerAspValArgProAsn----- 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 CGCTTCAAGGAGGACCTCTCGAACAATGGCCAGCGGTAAACATACGAAGAAGAATG 188

Qy 899 ----- 899

Db 189 GCTCGTCGACTGCTAGGTGCTGACAGTCGCAACTGCTTTTAATATTTCAGGAGCCAGAGAG 248

Qy 900 -----GlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIle 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 GAAACAGCTAATCAGGAATACAAAGTCTCCAGCTGTGCAACAGAGACTCATCAGTGAATA 308

Qy 917 GluPheArgLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAspGlu 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 GAGTACAGGCTAGAAAGGTCTCTCTGTGGATGAATCAGGTGATGAAGTTCAGTATGGAGAT 368

Qy 937 IleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgVal 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 GTGCCTGTGNAATGAATGGACCACTTTTCAGATGAAGCTGAACATCATTACAAGATC 428

Qy 957 ThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysVal 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 TTTGAGGGAATGCGCAGTAACTTTTCACTGTAGAGTGTGGAAATCCAAAGCCCAAGATC 488

Qy 977 TyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArg 996
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 TATTGGTTTAAAGATGGGAAGCAGATCTCTCCAAGAGTAGTATCACTACACATTTCAAAGA 548

Qy 997 GluGlyAspGlyThrCysSerLeuHisIleGluSerThrSerAspAspGlyAsn 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GATCTCGATGGACCTGCTCCCTCCATACCACAGCCCTACCTAGATGATGATGGGAAT 608

Qy 1017 TyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLysMetVal 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 TATACAAATATGCTGCAAAACCTCTCAGGGCGCATCAGTTGTACTGGACGGCTAATGTA 668

Qy 1037 GlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis----- 1052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 CAGGCTGTCAACCAAGAGGTGGAAGTCCCGGGTCTCCCTCAGGCGCATCTCTCAGC 728

Qy 1053 -----ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhe 1070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 729 AGGCCTCGTTCAGTCAAGGACAGTGGAGACCAAAATGAACCAATTCAGGACGGATTC 788

Qy 1071 PheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCys 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 789 TTCAGACCTCACTTCTTTCAGGCTCCTGGAGATCTGACTGTTCAAGAAAGGAAAACTCTGC 848

Qy 1091 ArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGly 1110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 AGAATGAGCTGCAAGTCAGTGGGTTCACCAACCCGATCTAAAGCTGAGCAACTAGATGGA 908

Qy 1111 GlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 909 AAGCCCGTACGCCCTGACAGTGCTCACAGATGCTGGTGGTGAACGGGGTGCACCTCT 968

Qy 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 969 CTGATCATAGAGCCAGTCACGTCACGTGATGCCGCGCATCTACACATGTATAGCTACCAAC 1028

Qy 1151 LysThrGlyGlnAsnSerPheSerLeuLeuSerValValAlaLysGluValLysLys 1170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1029 CGAGCAGGACAGAACTCATTCAGCCTGGAGCTTGTGGTCTGCTAAAGAAAGCACAAA 1088

```

```

QY 1171 AlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArg 1190
Db 1089 CCCCTGTGTTATGAGAGCTCCAAAACACAGGAGTCTCTGATGGTACCCAGTGGCG 1148
QY 1191 LeuGluCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGlu 1210
Db 1149 CTGGAATGCTGTATGGGAGTGCCACCACCTCAGATATTTTGGAGAAAGAAATGAA 1208
QY 1211 ThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCys 1230
Db 1209 TCACCTACTACAGCACTGACCGAGTGAGCATGCACACGAGCAACACGCTACATCTGC 1268
QY 1231 LeuLeuIleGlnProLysLysSerAspAlaGlyTyrTrpThrLeuSerAlaLysAsn 1250
Db 1269 CTGCTATTTCAGGAGGCCACAAAAGAGATGCTGGGTGATATCTGTGTCAGCAAGAAT 1328
QY 1251 GluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis--- 1269
Db 1329 GAAGCAGGAGTGTGCTGTACTGCGAGCTGGAGCTTTTACACCCAGTGGCATCAGCAG 1388
QY 1270 ---GlnIleProProMetSerValArgProSerGlySerArgTyrGlySerLeuThr 1288
Db 1389 TCACAGACGACCAAGCAAAAAGTACGGCCCTCAGCCAGTGCCTATGCGACACTTCG 1448
QY 1289 SerLysGlyLeuAspIlePheSerAlaPheSerMetGluSerThrMetValTyrSer 1308
Db 1449 GACCAGGAGTACATCAAGCAGCGTTCACACCT---GAGGCCAACCCATCTCACCTG 1505
QY 1309 CysSerSerArgSerValValGluSerAspGluLeu 1320
Db 1506 ACACCTGAATACTGCTTGGTGAAGTGTGAGGACCTG 1541

RESULT 15
AF205078
LOCUS
DEFINITION
Mus musculus strain Swiss Webster/NIH actin-associated protein
palladin mRNA, partial cds.
ACCESSION
AF205078
VERSION
AF205078.1 GI:9828172
KEYWORDS
Mus musculus.
SOURCE
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1580)
Parast.M.M. and Otey.C.A.
Characterization of palladin, a novel protein localized to stress
fibers and cell adhesions
J. Cell Biol. 150 (3), 643-656 (2000)
20391984
PUBMED
10931874
REFERENCE
2 (bases 1 to 1580)
Parast,M. and Otey,C.
Direct Submission
Submitted (16-NOV-1999) Department of Cell and Molecular
Physiology, University of North Carolina-Chapel Hill, 52 MSRB Bldg,
Campus Box 7545, Chapel Hill, NC 27599, USA
JOURNAL
Location/Qualifiers
FEATURES
1..1580
/organism="Mus musculus"
/strain="Swiss Webster/NIH"
/db_xref="taxon:10090"
/dev_stage="11.5 day embryo"
<1..1479
/note="Involved in cytoskeletal assembly"
/codon_start=1
/product="actin-associated protein palladin"
/protein_id="AG00078.1"
/db_xref="GI:9828173"
/translation="ASDEIQTGKDAVTDLKERLRFKEDLLNNGSPRLTYBERMARR
LIGADSANVFTQEEFEATNAGYKVSCEQLRISEIYRLERSPVDESGDVQPDV
PVENATAPFEEKLKHKIFEGMPVFTFCRVAGNPKPKTYFKDKGKQISPKSDHYTIQ
RDLDTGCSLHTTASTLLDDNGNTIMAA NPQGRVSVCTGRLMVQAVNQRGRSPSPSGHP
source
CDS

```

```

HARRPRSRSDSDENEPIQERFRPHFLOAPCDLTVOEGKLCRMDCKVSGLPTDLS
WOLDGKPIRPSAHLKMLRENGVHSLIIEPVTSRDAGIVTCTATNRAGONSFNLELV
AAKEAHKAPVEKELQNTGVADGYPVRLKESVSGVPPPOIEMKKNESJTHSTERSVM
HODNHYICLLIQGATKEDAGWYTSKNEAGIVSCTARLDYVTOHQHQPQTTPKKV
RFSASRYAALSDQGLDIKAAFOPEASPSHLTLNGLVSEDL"
1480..>1580
3'UTR
BASE COUNT 415 a 437 c 427 g 301 t
ORIGIN
Alignment Scores:
Pred. No.: 1,44e-47 Length: 1580
Score: 1460.00 Matches: 274
Percent Similarity: 71.54% Conservative: 78
Best Local Similarity: 55.69% Mismatches: 96
Query Match: 21.19% Indels: 44
DB: 10 Gaps: 5
US-09-818-990B-2 (1-1320) x AF205078 (1-1580)
QY 872 AsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIle 891
Db 4 TCTGATGAGGAGATTCAAGGCACAAAGGATGCTGTCTATCCAAAGACCTGGCAACGAGCTT 63
QY 892 ThrPheSerSerValArgProAsn----- 899
Db 64 CGCTTCAAGGAGGACCTTCTGAACAATGGCCAACCGAGGCTAACTATGAGAAAGAATG 123
QY 899 ----- 899
Db 124 GCTCGCCGCTGCTTGGAGCCGACAGCGCAACAGCTCTTCAACATCCAGAGCCAGAGAA 183
QY 900 -----GlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIle 916
Db 184 ACGGACGCCAATCAGGAGTACAAGTCTCTAGCTGCGAGCAGAGGCTGATTAGCGAGATT 243
QY 917 GluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAspGlu 936
Db 244 GAGTACCGGCTGGAGCGCTCCCTGTGTGATGAGTGGGAGACGAGGTGAGGATCCAGAT 303
QY 937 IleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgVal 956
Db 304 GTCCCTGTGGAGAACGCAACAGCTCCCTCTTTGAGATGAAGTGAACACTACAAGAT 363
QY 957 ThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysVal 976
Db 364 TTTGAGGGGATCGCGTGACTTTCACGTGCGAGTGGCTGGGATCCAAAGCCAAAGATC 423
QY 977 TyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArg 996
Db 424 TATTGGTTTAAAGATGGGAAGCAGATTCTCCGAAGAGCGATCACTACACCATTCAGAGA 483
QY 997 GluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspAspGlyAsn 1016
Db 484 GACCTTGTGGACCTCTCTCTCCACACACCGCCCTCTACCCCTAGACGATGGGAAC 543
QY 1017 TyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetVal 1036
Db 544 TACACCATCATGGCTGCCAACCCCTCAGGGTGGCTGAGTCAAGTGTACAGGAAGGCTAATGGTA 603
QY 1037 GlnSerLeuProIleArgSerArg-----LeuThrSerAlaGlyGlnSerHisArgGly 1054
Db 604 CAGGCTGTCAACCAAGAGGCGCGAGCTCCCGCTCTCCCTCAGGCCATCTCATGCCAGA 663
QY 1055 ArgSerArgValGlnGluArgAsp-----LysGluProLeuGlnGluArgPhe 1070
Db 664 AGGCTCGCTCTCGATCACGGAGTGGAGATGAAAACGAGGCCATTCAGGAGCGATT 723
QY 1071 PheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCys 1090
Db 724 TTCAGACCTCACCTCCCTGAGGCTCTCGAGACCTGACCGGTTTCAGGAAGGCGAGCTCTCG 783
QY 1091 ArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGly 1110

```

```
Db 784 AGGATGGACTGCCAAGGTCAGTGGGTTACCAACCCAGATCTCAGCTGGCAACTAGATCGA 843
Qy 1111 GlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisser 1130
    ::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 844 AAGCCCATACCCCGCAGCAGTGTCTACAAGATGGTCCGTGAGAAATGGGGTCCACTCC 903
Qy 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrlsCysIleAlaThrAsn 1150
    ::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 904 CTCATTATAGAGCAGTCACGTCCTCCGGGAGCCCGGCATCTACACATGTATTGCCACCAAC 963
Qy 1151 LysThrGlyGlnAnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys 1170
    ::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 964 AGACGAGACAGAACTCGTTTAACCTGGAGCTTGTCTGCTAAGGAAGACACAAAG 1023
Qy 1171 AlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArg 1190
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1024 GCCCTGTGTTTATGGAGAACTACAGAACACGGGGGTGCTGATGATACCCAGTGGG 1083
Qy 1191 LeuGluCysArgValIleGlyMetProProValPheTyrlsLysAspAsnGlu 1210
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1084 CTGGAATGCCGTGTCGGGAGTCCGCCACCTCAGATATTTTGGAAAGAAATAATGAA 1143
Qy 1211 ThrIleProCysThrArgGluArgIleSerMethisGlnaspThrThrGlyTyrlsAlaCys 1230
    ::::: ::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1144 TCGCTCACTCACAGCACTGAGCGAGTAAGCATGCACGAGGATAATCATGGCTACATCTGC 1203
Qy 1231 LeuLeuIleGlnProAlaLysLysSerAspAlaGlyTyrlsThrLeuSerAlaLysAsn 1250
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1204 CTGCTCATCCAGGAGGCCACAAAGAGAGACCTGGGTGGGTACACTGTGTCGCCCAAGAAC 1263
Qy 1251 GluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrlsAlaGlnTrpHisHis-- 1269
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1264 GAAGCAGGCATTGTCTCTGCACTGCCCGCTGGATGCTACACCCAGTGGCACCAGCAG 1323
Qy 1270 ---GlnIleProProMetSerValArgProSerGlySerArgTyrlsGlySerLeuThr 1288
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1324 CCACAGACCACCAAGCCAAAGAGTACGGCCCTCGGCCACTCGCTACGCGACACTTTCG 1383
Qy 1289 SerLysGlyLeuAspIlePheSerAlaPheSerSerMetGluSerThrMetValTyrlsSer 1308
    ::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1384 GACCAGGAGCTAGACATCAAGCCGCTTTCACACT---GAAGCCAGCCCATCTCACCTG 1440
Qy 1309 CysSerArgSerValValGluSerAspGluLeu 1320
    ::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1441 ACTCTGAACACCGGCTTGGTAGAAAGTGAAGACCTG 1476
```

Search completed: November 30, 2002, 20:58:57
Job time : 6041 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: November 30, 2002, 17:03:50 ; Search time 73 Seconds
(without alignments)
14467.732 Million cell updates/sec
Title: US-09-818-990B-1
Perfect score: 1316
Sequence: 1 atcgacgacagcatag.....tggagagtgtgaactttaa 3963
Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 908470 seqs, 133250620 residues
Word size: 1
Total number of hits satisfying chosen parameters: 1687582
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 1000 summaries
Command line parameters: -DEV=xlh
-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09818990/runat_26112002_093221_22022/app_query.fasta_1.4103
-DB=A_Geneseq_101002 -QPMT=fastan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09818990_@C9N_1_1_91-runat_26112002_093221_22022 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	494	37.5	507	22	Human protein sequ
2	261	19.8	264	22	Human cDNA SEQ ID
3	225	17.1	408	22	Novel human diago
4	179	13.6	261	22	Human immunoglobul
5	113	8.6	159	22	Human cDNA SEQ ID
6	113	8.6	159	22	Human immunoglobul
7	81	6.2	130	21	Human ORFX ORF639
8	33	2.5	50	23	Human ORFX protein
9	19	1.4	433	21	Human ORFX ORF2526
10	19	1.4	544	21	Human colon cancer
11	16	1.2	43	22	Human secreted pro
12	16	1.2	44	22	Human secreted pro
13	16	1.2	49	22	Human secreted pro
14	14	1.1	104	22	Human foetal prote
15	14	1.1	162	23	Human ORFX protein
16	11	0.8	776	22	Novel human diago
17	11	0.8	776	22	Human anglogenesis
18	11	0.8	776	23	Human PRO4527 prot
19	11	0.8	777	22	Novel human diago
20	10	0.8	881	22	Novel human diago
21	10	0.8	881	22	Novel human diago
22	10	0.8	1892	21	Arabidopsis thalia
23	10	0.8	1940	21	Arabidopsis thalia
24	10	0.8	1944	21	Arabidopsis thalia
25	9	0.7	223	23	Human DTHP polype
c 26	9	0.7	3502	22	Drosophila melanog
c 27	8	0.6	14	21	HHV8 ORF 65-deri
28	8	0.6	17	22	Peptide #9818 enco
29	8	0.6	17	22	Human brain expres
30	8	0.6	17	22	Human bone marrow
31	8	0.6	17	22	Peptide #10157 enc
32	8	0.6	31	21	Human secreted pro
33	8	0.6	31	22	muO-conopeptide G2
c 34	8	0.6	39	22	Human secreted pro
c 35	8	0.6	46	22	Peptide #5789 enco
c 36	8	0.6	46	22	Protein #5463 enco
c 37	8	0.6	46	22	Human brain expres
c 38	8	0.6	46	22	Human bone marrow
c 39	8	0.6	46	22	Peptide #5511 enco
c 40	8	0.6	46	22	Peptide #5753 enco
41	8	0.6	46	22	Human secreted pro
c 42	8	0.6	46	23	Human peptide enco
c 43	8	0.6	48	22	Drosophila melanog
44	8	0.6	51	22	Peptide #6710 enco
45	8	0.6	51	22	Protein #6058 enco
46	8	0.6	51	22	Human brain expres
47	8	0.6	51	22	Human bone marrow
48	8	0.6	51	22	Peptide #6022 enco
49	8	0.6	51	22	Peptide #6739 enco
50	8	0.6	51	23	Human peptide enco
51	8	0.6	52	22	Propionibacterium
52	8	0.6	58	22	Human secreted pro
53	8	0.6	63	22	Human colon cancer
54	8	0.6	66	22	Propionibacterium
55	8	0.6	74	22	Human nervous syst
c 56	8	0.6	77	22	Propionibacterium
c 57	8	0.6	80	22	muO-conopeptide G2
58	8	0.6	85	22	Propionibacterium
59	8	0.6	91	23	Human ORFX protein
c 60	8	0.6	93	20	Human 5' EST secre
c 61	8	0.6	93	21	Human 5' EST relat
c 62	8	0.6	98	21	Arabidopsis thalia
63	8	0.6	108	20	B. burgdorferi ant
64	8	0.6	111	22	Propionibacterium
c 65	8	0.6	114	22	Human kinase homol
66	8	0.6	119	22	Propionibacterium
c 67	8	0.6	119	22	Propionibacterium
68	8	0.6	129	20	B. burgdorferi ant

69	8	0.6	133	22	AA093627	Human polypeptide,	142	8	0.6	362	23	AB092431	Herbicidally activ
c 70	8	0.6	134	23	ABP07102	Human ORFX protein	c 143	8	0.6	368	21	AB026843	Sakuranetin synth
c 71	8	0.6	141	20	AA036045	Extended human sec	144	8	0.6	381	21	AA039847	Arabidopsis thalia
72	8	0.6	142	21	AA093408	Amino acid sequenc	c 145	8	0.6	389	21	AA036220	Murine beta-1,3-ga
73	8	0.6	146	23	ABP42439	Human ovarian anti	c 146	8	0.6	392	12	AA014662	El-beta subunit of
c 74	8	0.6	149	22	AA060529	Propionibacterium	c 147	8	0.6	397	21	AA036619	Human beta-1,3-gal
75	8	0.6	158	21	AA055518	Arabidopsis thalia	c 148	8	0.6	437	22	AA082933	S. epidermidis ope
76	8	0.6	165	21	AA066433	Arabidopsis thalia	149	8	0.6	464	22	AB065760	Drosophila melanog
77	8	0.6	173	21	AA055517	Arabidopsis thalia	150	8	0.6	501	19	AA082571	Human Bp1 DNA. H
78	8	0.6	173	22	AA038789	Novel subtilisin h	151	8	0.6	519	23	AA093044	Arabidopsis transc
79	8	0.6	173	22	AA038805	Novel subtilisin h	c 152	8	0.6	595	22	AA036514	Pseudomonas aerugi
80	8	0.6	173	22	AA038859	Novel subtilisin h	c 153	8	0.6	603	22	AB064542	Drosophila melanog
81	8	0.6	174	22	AB014566	Novel human diagno	c 154	8	0.6	605	23	ABP39580	Staphylococcus epi
82	8	0.6	176	21	AA093412	Amino acid sequenc	155	8	0.6	646	21	AA057307	P. aeruginosa matu
c 83	8	0.6	178	22	AA012073	Dendritic cell (DC	156	8	0.6	670	21	AA057314	P. aeruginosa cera
c 84	8	0.6	179	22	AB021767	Novel human diagno	157	8	0.6	745	23	AB057319	Mouse ischaemic co
c 85	8	0.6	192	19	AA074885	Human secreted pro	158	8	0.6	746	20	AA028882	Human Ext-1 protei
c 86	8	0.6	192	21	AA091681	Human secreted pro	159	8	0.6	748	23	AB092874	Herbicidally activ
c 87	8	0.6	198	23	AB048986	Listeria monocytog	160	8	0.6	761	22	AB058572	Drosophila melanog
c 88	8	0.6	211	21	AA023517	Arabidopsis thalia	161	8	0.6	783	22	AB049033	Novel human diagno
89	8	0.6	217	21	AA066432	Arabidopsis thalia	162	8	0.6	792	22	AB064798	Drosophila melanog
90	8	0.6	218	21	AA024454	Human secreted pro	c 163	8	0.6	793	22	AB065720	A. gossypii AG011
91	8	0.6	219	21	AA023516	Arabidopsis thalia	164	8	0.6	848	21	AA070955	Arabidopsis SERK 1
c 92	8	0.6	221	22	AA045702	Human 7TM clone HD	c 165	8	0.6	996	22	AA090093	C glutamicum prote
c 93	8	0.6	225	21	AA094449	Human inflammation	166	8	0.6	1051	12	AA014117	Hamster GAP b3 pro
c 94	8	0.6	225	22	AA065272	Human cell surface	167	8	0.6	1250	21	AA091279	Group B Streptococ
c 95	8	0.6	229	21	AA066431	Arabidopsis thalia	168	8	0.6	1252	23	ABP26711	Streptococcus poly
96	8	0.6	229	21	AA039849	Arabidopsis thalia	169	8	0.6	1252	23	ABP29749	Streptococcus poly
c 97	8	0.6	239	22	AA074584	Human HAERBs isof	c 170	8	0.6	1285	21	AA043359	Human ORFX ORF3123
c 98	8	0.6	240	21	AA026845	Sakuranetin synth	171	8	0.6	1337	16	AA085203	huDP-1. Homo sap
c 99	8	0.6	242	21	AA073495	Human secreted pro	172	8	0.6	1510	20	AA013112	Glutamine-oxogluta
c 100	8	0.6	248	20	AA015225	Human receptor pro	173	8	0.6	1510	22	AA089953	C glutamicum prote
c 101	8	0.6	248	20	AA036046	Extended human sec	174	8	0.6	1510	22	AA079669	Corynebacterium gl
c 102	8	0.6	248	21	AA091531	Human secreted pro	175	8	0.6	1527	20	AA043543	A human MPR-relate
c 103	8	0.6	248	22	AA012072	Dendritic cell (DC	176	8	0.6	1527	23	AA091309	Human protein NOV1
c 104	8	0.6	248	22	AA070469	Human HAERBs-iso	177	8	0.6	1528	18	AA033363	Human multidrug re
c 105	8	0.6	248	23	AB090341	Human polypeptide	178	8	0.6	1953	23	AA084351	Protein MYLK diffe
c 106	8	0.6	249	22	AA074583	Human HAERBs isof	c 179	8	0.6	2023	22	AB063487	Drosophila melanog
c 107	8	0.6	250	20	AA048505	Human breast tumor	c 180	8	0.6	2053	22	AA035011	Human protein kina
c 108	8	0.6	257	21	AA058419	Lung cancer associ	c 181	8	0.6	2053	23	AAE24079	Human MPR protein
c 109	8	0.6	259	22	AA036373	Gopher apple oleoy	c 182	8	0.6	2054	23	AAE24150	Human kinase (PKIN
110	8	0.6	271	21	AA014830	Bacillus lentus I-	c 183	8	0.6	2547	19	AA081511	Drosophila fat fac
c 111	8	0.6	273	21	AA091680	Human secreted pro	c 184	8	0.6	2559	23	AB057181	Mouse ischaemic co
c 112	8	0.6	273	22	AB011989	Human secreted pro	c 185	8	0.6	2587	22	ABG20102	Novel human diagno
c 113	8	0.6	273	22	AA025809	Human protein sequ	c 186	8	0.6	2644	22	ABG20103	Novel human diagno
c 114	8	0.6	273	22	AA075619	Human colon cancer	c 187	8	0.6	2778	22	AB058683	Drosophila melanog
c 115	8	0.6	279	22	AA093492	Human polypeptide,	c 188	8	0.6	3594	23	AAE20147	Mouse C3b/C4b comp
c 116	8	0.6	295	23	AB048857	Listeria monocytog	189	7	0.5	9	17	AA000714	Human Stat 5 fragm
c 117	8	0.6	298	22	AA041987	Human polypeptide	190	7	0.5	10	20	AA045955	Immunogenic peptid
c 118	8	0.6	299	22	ABG09593	Novel human diagno	c 191	7	0.5	10	22	AB071919	Random 10-mer pept
c 119	8	0.6	305	23	AA016509	Rice caffeic acid	192	7	0.5	10	23	ABP30935	HLA-A2 10-mer bind
120	8	0.6	309	21	AA023515	Arabidopsis thalia	c 193	7	0.5	11	21	AB090394	Hepatitis CB virus
c 121	8	0.6	335	22	ABG06645	Novel human diagno	c 194	7	0.5	14	22	AA068142	Peptide derived fr
122	8	0.6	339	21	AA054254	Human pancreatic c	c 195	7	0.5	14	22	AA068143	Peptide derived fr
123	8	0.6	352	18	AA014795	Mangosteen Class I	c 196	7	0.5	14	22	AA068174	Peptide derived fr
124	8	0.6	352	18	AA009298	Mangosteen Class I	c 197	7	0.5	14	22	AA068175	Peptide derived fr
125	8	0.6	352	20	AA087471	Mangosteen acyl AC	c 198	7	0.5	14	22	AA068176	Peptide derived fr
126	8	0.6	352	20	AA087472	Mangosteen acyl AC	c 199	7	0.5	14	22	AA068177	Peptide derived fr
127	8	0.6	352	20	AA098150	Mangosteen acyl AC	200	7	0.5	15	18	AA038981	Peptide resembling
128	8	0.6	352	20	AA087469	Mangosteen acyl AC	201	7	0.5	15	22	AA068653	Human cytomagalovi
129	8	0.6	352	20	AA087466	Mangosteen C18:1 a	202	7	0.5	15	22	AA068654	Human cytomagalovi
130	8	0.6	352	21	AA082090	Mangosteen FataA-ty	203	7	0.5	15	22	AA068655	Human cytomagalovi
131	8	0.6	352	23	AA096742	Mangosteen FataA-ty	204	7	0.5	18	22	AA089036	HIV gp120 protein
132	8	0.6	356	22	AA025563	Human G Protein-Co	205	7	0.5	18	22	AA089037	HIV gp120 protein
133	8	0.6	362	13	AA029172	Brassica thioester	206	7	0.5	18	22	AA089038	HIV gp120 protein
134	8	0.6	362	15	AA034948	Brassica campestri	207	7	0.5	20	20	AA039286	G-protein coupled
135	8	0.6	362	16	AA074150	Brassica acyl thio	c 208	7	0.5	21	18	AA034050	Human MDR1-p glyco
136	8	0.6	362	19	AA044333	Brassica campestri	209	7	0.5	21	18	AA034004	Fusin receptor ant
137	8	0.6	362	20	AA028638	Brassica napus acy	c 210	7	0.5	21	22	AA04308	ATP-binding cass
138	8	0.6	362	20	AA074667	Brassica rapa C18:	211	7	0.5	21	23	ABP30910	OE8 antibody epit
139	8	0.6	362	21	AA039848	Arabidopsis thalia	212	7	0.5	21	23	AA089534	Insulin/insulin-li
140	8	0.6	362	21	AA053760	An acyl-ACP thioes	213	7	0.5	23	20	AA039241	G-protein coupled
141	8	0.6	362	21	AA053761	An acyl-ACP thioes	214	7	0.5	24	22	AA060446	Human brain expres

c 215	7	0.5	24	22	AAW73087	Human bone marrow	c 288	7	0.5	52	22	ABB29542	Peptide #2193 enco
c 216	7	0.5	24	22	AAW33306	Peptide #7343 enco	c 289	7	0.5	52	22	ABB34727	Peptide #2233 enco
c 217	7	0.5	24	22	ABG42933	Human peptide enco	c 290	7	0.5	52	22	ABB20139	Protein #2138 enco
c 218	7	0.5	25	22	ABBA0984	Peptide #8490 enco	c 291	7	0.5	52	22	AAW81029	Human haematologic
c 219	7	0.5	25	22	AAW61844	Human brain expres	c 292	7	0.5	52	22	AAW81262	Human haematologic
c 220	7	0.5	25	22	AAW34759	Peptide #8796 enco	c 293	7	0.5	52	22	AAW55524	Human brain expres
c 221	7	0.5	26	10	AAW91262	Inhibin 18 kD chai	c 294	7	0.5	52	22	AAW67908	Human bone marrow
c 222	7	0.5	26	23	AAW93966	Bovine inhibin alp	c 295	7	0.5	52	22	AAW15725	Peptide #2159 enco
c 223	7	0.5	28	22	AAW51686	PAM related peptid	c 296	7	0.5	52	22	AAW28230	Peptide #2267 enco
c 224	7	0.5	29	9	AAW81907	N-terminal of inhi	c 297	7	0.5	52	22	AAW03459	Peptide #2141 enco
c 225	7	0.5	30	22	ABW37575	Peptide #5081 enco	c 298	7	0.5	52	23	ABG37444	Human peptide enco
c 226	7	0.5	30	22	AAW89359	Human immune/haema	c 299	7	0.5	52	23	ABP10517	Human ORFX protein
c 227	7	0.5	32	22	AAW92074	Inhibin peptide SE	c 300	7	0.5	53	22	ABG22991	Novel human diagno
c 228	7	0.5	32	22	AAW92078	Inhibin peptide SE	c 301	7	0.5	53	22	ABW29253	Peptide #1904 enco
c 229	7	0.5	33	12	AAW12087	N-terminal of 18 k	c 302	7	0.5	53	22	ABW34418	Peptide #1924 enco
c 230	7	0.5	33	22	AAW92075	Inhibin peptide SE	c 303	7	0.5	53	22	ABW19830	Protein #1829 enco
c 231	7	0.5	33	22	AAW92079	Inhibin peptide SE	c 304	7	0.5	53	22	AAW52508	Human brain expres
c 232	7	0.5	34	15	AAW66852	Rice mitochondrial	c 305	7	0.5	53	22	AAW67603	Human bone marrow
c 233	7	0.5	34	15	AAW74754	Rice mitochondrial	c 306	7	0.5	53	22	AAW15410	Peptide #1844 enco
c 234	7	0.5	34	21	AAW28738	Human secreted pro	c 307	7	0.5	53	22	AAW27899	Peptide #1936 enco
c 235	7	0.5	34	21	ABG66875	Human prostate spe	c 308	7	0.5	53	22	AAW03172	Peptide #1854 enco
c 236	7	0.5	35	21	AAW45245	Human secreted pro	c 309	7	0.5	53	22	ABG37193	Human peptide enco
c 237	7	0.5	35	21	AAW56759	Arabidopsis thalia	c 310	7	0.5	54	22	AAW59008	Propionibacterium
c 238	7	0.5	36	22	ABG14004	Novel human diagno	c 311	7	0.5	54	22	AAW67102	Propionibacterium
c 239	7	0.5	37	21	AAW13519	Human glycine tran	c 312	7	0.5	54	22	AAU14366	Human novel protei
c 240	7	0.5	37	21	AAW56758	Arabidopsis thalia	c 313	7	0.5	55	20	AAW19563	Amino acid sequenc
c 241	7	0.5	38	22	AAW89067	Human immune/haema	c 314	7	0.5	55	21	ABG35586	Arabidopsis thalia
c 242	7	0.5	39	20	AAW74414	HPMBO91 protein se	c 315	7	0.5	55	22	ABG16736	Novel human diagno
c 243	7	0.5	39	22	ABW44117	Peptide #11623 enc	c 316	7	0.5	56	22	AAU41382	Propionibacterium
c 244	7	0.5	39	22	ABW27004	Peptide #9003 enco	c 317	7	0.5	56	23	ABP32402	Human ORF1375 prot
c 245	7	0.5	39	22	AAW65137	Human brain expres	c 318	7	0.5	57	21	AAW57357	Arabidopsis thalia
c 246	7	0.5	39	22	AAW77844	Human bone marrow	c 319	7	0.5	57	21	AAW60512	Arabidopsis thalia
c 247	7	0.5	39	22	AAW21748	Peptide #8182 enco	c 320	7	0.5	57	23	ABP01126	Human ORFX protein
c 248	7	0.5	39	22	AAW38084	Peptide #12101 enc	c 321	7	0.5	58	22	AAU54686	Propionibacterium
c 249	7	0.5	39	23	ABG46872	Human peptide enco	c 322	7	0.5	58	22	AAU55851	Propionibacterium
c 250	7	0.5	40	23	AAW29090	Streptococcus poly	c 323	7	0.5	58	22	ABG28606	Novel human diagno
c 251	7	0.5	41	13	AAW25795	[GLu2,13,22, Leu12	c 324	7	0.5	58	22	AAW75464	Human colon cancer
c 252	7	0.5	42	20	AAW36254	Human secreted pro	c 325	7	0.5	58	22	AAU00689	Thymosin-beta-10-1
c 253	7	0.5	42	21	AAW56198	Human secreted pro	c 326	7	0.5	58	23	AAW23797	Thymosin-beta-10-1
c 254	7	0.5	42	22	ABW41833	Peptide #9339 enco	c 327	7	0.5	59	22	AAW45252	Propionibacterium
c 255	7	0.5	42	22	AAW62706	Human brain expres	c 328	7	0.5	59	22	AAU50067	Propionibacterium
c 256	7	0.5	42	22	AAW75523	Human bone marrow	c 329	7	0.5	59	22	AAU30424	Novel human secret
c 257	7	0.5	42	22	AAW35630	Peptide #9667 enco	c 330	7	0.5	60	22	AAW67556	Propionibacterium
c 258	7	0.5	42	22	AAW65088	Gene #19 associate	c 331	7	0.5	60	22	AAW88937	Human immune/haema
c 259	7	0.5	42	23	ABW45074	Human peptide enco	c 332	7	0.5	60	22	AAW91442	Human immune/haema
c 260	7	0.5	44	20	AAW12028	Human 5' EST secre	c 333	7	0.5	61	22	AAW60201	Propionibacterium
c 261	7	0.5	44	22	ABG30028	Novel human diagno	c 334	7	0.5	61	22	AAW62142	Propionibacterium
c 262	7	0.5	46	22	ABG05181	Novel human diagno	c 335	7	0.5	61	22	AAW67180	Propionibacterium
c 263	7	0.5	47	22	ABW41064	Peptide #8570 enco	c 336	7	0.5	61	22	ABW42326	Peptide #9832 enco
c 264	7	0.5	47	22	ABW17487	Human nervous syst	c 337	7	0.5	61	22	ABW25817	Protein #7816 enco
c 265	7	0.5	47	22	AAW61922	Human brain expres	c 338	7	0.5	61	22	AAW63211	Human brain expres
c 266	7	0.5	47	22	AAW74724	Human bone marrow	c 339	7	0.5	61	22	AAW76025	Human bone marrow
c 267	7	0.5	47	22	AAW34841	Peptide #8878 enco	c 340	7	0.5	61	22	AAW20772	Peptide #7206 enco
c 268	7	0.5	48	20	AAW02741	Human secreted pro	c 341	7	0.5	61	22	AAW24067	Human EST encoded
c 269	7	0.5	48	22	ABW43462	Peptide #10968 enc	c 342	7	0.5	61	22	AAW36131	Peptide #10168 enc
c 270	7	0.5	48	22	ABW26428	Protein #8427 enco	c 343	7	0.5	61	22	AAW90811	Human shear stress
c 271	7	0.5	48	22	AAW64392	Human brain expres	c 344	7	0.5	61	22	ABW45423	Human peptide enco
c 272	7	0.5	48	22	AAW77212	Human bone marrow	c 345	7	0.5	62	22	AAU50072	Propionibacterium
c 273	7	0.5	48	22	AAW21145	Peptide #7579 enco	c 346	7	0.5	62	22	ABW01370	Novel human diagno
c 274	7	0.5	48	22	AAW37350	Peptide #11387 enc	c 347	7	0.5	63	23	ABW41270	Human ovarian anti
c 275	7	0.5	48	23	ABW46225	Human peptide enco	c 348	7	0.5	64	20	AAW88672	Secreted protein e
c 276	7	0.5	49	22	ABW10598	Human pancreatic c	c 349	7	0.5	64	21	AAW59891	Arabidopsis thalia
c 277	7	0.5	49	22	AAW92351	Human digestive sy	c 350	7	0.5	64	22	ABG10533	Novel human diagno
c 278	7	0.5	50	22	AAU44491	Propionibacterium	c 351	7	0.5	64	22	ABW50439	Human secreted pro
c 279	7	0.5	50	22	ABW50550	Human secreted pro	c 352	7	0.5	65	22	AAW92183	C glutamicum prote
c 280	7	0.5	50	22	AAW00968	Human bone marrow	c 353	7	0.5	66	21	AAW56342	Human secreted pro
c 281	7	0.5	50	22	ABW67768	Fragment from a wh	c 354	7	0.5	66	21	AAW32992	Pinus radiata tran
c 282	7	0.5	50	23	ABW07947	Human ORFX protein	c 355	7	0.5	66	21	AAW38527	Human immune/haema
c 283	7	0.5	51	21	AAW41803	Human ORFX ORF1567	c 356	7	0.5	66	22	AAW82678	Human secreted pro
c 284	7	0.5	51	22	AAW48303	Propionibacterium	c 357	7	0.5	66	22	AAW90726	C glutamicum prote
c 285	7	0.5	51	22	ABW42254	Peptide #9760 enco	c 358	7	0.5	67	22	ABW28031	Novel human diagno
c 286	7	0.5	51	22	AAW75954	Human bone marrow	c 359	7	0.5	68	21	AAW56982	Human prostate can
c 287	7	0.5	52	22	AAW58684	Propionibacterium	c 360	7	0.5	68	21	AAW19652	Arabidopsis thalia

361	7	0.5	68	22	AAU44247	Propionibacterium	c 434	7	0.5	91	22	AAAG75213	Human colon cancer
362	7	0.5	68	22	AAU53359	Propionibacterium	c 435	7	0.5	92	22	ABBI5424	Human nervous syst
c 363	7	0.5	68	22	AAU84261	Human immune/haema	436	7	0.5	93	21	AAB42262	Human ORFX ORF2026
364	7	0.5	68	22	AAU91282	Human immune/haema	437	7	0.5	93	22	ABB37831	Peptide #5337 enco
c 365	7	0.5	68	23	ABP34368	Human ORF3341 prot	438	7	0.5	93	22	ABB23097	Protein #5036 enco
366	7	0.5	68	23	ABP34381	Human ORF3354 prot	439	7	0.5	93	22	AAM58451	Human brain expres
c 367	7	0.5	69	22	AAU48937	Propionibacterium	440	7	0.5	93	22	AAU70943	Human bone marrow
c 368	7	0.5	69	22	AAU57779	Propionibacterium	441	7	0.5	93	22	AAM89523	Human immune/haema
369	7	0.5	69	23	ABP09342	Human ORFX protein	442	7	0.5	93	22	AAO03827	Human polypeptide
370	7	0.5	70	20	AAU13154	Human secreted pro	443	7	0.5	93	22	AAU31233	Peptide #5270 enco
c 371	7	0.5	71	21	AAG11076	Arabidopsis thalia	444	7	0.5	93	22	AAU21634	Novel human neopla
c 372	7	0.5	71	21	AAG56917	Arabidopsis thalia	445	7	0.5	93	23	ABG40742	Human peptide enco
c 373	7	0.5	71	22	AAM85033	Human immune/haema	446	7	0.5	93	23	ABP08264	Human ORFX protein
c 374	7	0.5	71	22	AAO13297	Human polypeptide	447	7	0.5	94	21	AAB58873	Breast and ovarian
375	7	0.5	72	22	AAM89498	Human immune/haema	448	7	0.5	94	21	AAG57637	Arabidopsis thalia
376	7	0.5	72	22	AAE03555	Human trefoil doma	449	7	0.5	94	22	AAU63724	Propionibacterium
c 377	7	0.5	73	21	AAG44945	Zea mays protein f	450	7	0.5	94	23	AAU77166	Corticotrophin rel
c 378	7	0.5	74	18	AAG10031	Protein encoded by	c 451	7	0.5	95	21	AAB41343	Human ORFX ORF1107
c 379	7	0.5	74	22	ABBI10140	Human cDNA SEQ ID	452	7	0.5	95	22	AAU61984	Propionibacterium
c 380	7	0.5	74	22	AAU20962	Human novel foetal	c 453	7	0.5	95	22	AAU62534	Human immune/haema
c 381	7	0.5	74	22	AAE09685	Human gene 3 encod	c 454	7	0.5	95	23	ABP10444	Human ORFX protein
382	7	0.5	75	21	AAG23802	Arabidopsis thalia	c 455	7	0.5	95	23	ABP10444	Human secreted pro
c 383	7	0.5	75	21	AAG02356	Human secreted pro	c 456	7	0.5	96	23	ABP41307	Human ovarian anti
384	7	0.5	75	23	ABP32432	Human ORF1405 prot	c 457	7	0.5	97	21	AAB16757	Bacteriophage Dp-1
c 385	7	0.5	75	23	ABP33979	Human ORF2952 prot	458	7	0.5	97	22	AAU48843	Propionibacterium
c 386	7	0.5	75	23	ABP00942	Human ORFX protein	459	7	0.5	97	22	AAU55127	Propionibacterium
387	7	0.5	76	22	ABBI7866	Human nervous syst	460	7	0.5	98	22	AAU67006	Propionibacterium
388	7	0.5	77	21	AAG00855	Human secreted pro	461	7	0.5	98	22	AAU55100	Propionibacterium
c 389	7	0.5	77	23	ABP35279	Human ORF4252 prot	462	7	0.5	98	22	ABBI15982	Human nervous syst
c 390	7	0.5	77	23	ABP01001	Human ORFX protein	463	7	0.5	98	22	ABBI17004	Human nervous syst
391	7	0.5	78	21	AAB54296	Human pancreatic c	464	7	0.5	98	22	AAG76522	Human colon cancer
392	7	0.5	78	22	AAM86714	Human immune/haema	c 465	7	0.5	98	22	AAG89223	Human secreted pro
c 393	7	0.5	78	22	AAO05173	Human polypeptide	c 466	7	0.5	98	22	AAU00284	Immunogenic epitop
c 394	7	0.5	78	22	AAU06055	Cone snail O-supe	c 467	7	0.5	99	22	AAM88579	Human immune/haema
395	7	0.5	78	23	ABP77048	Human protein sequ	468	7	0.5	100	20	AAU48251	Human prostate can
396	7	0.5	78	23	ABG54933	Human albumin fusi	c 469	7	0.5	100	22	AAU55373	Propionibacterium
397	7	0.5	79	22	AAU48480	Propionibacterium	470	7	0.5	100	22	AAG70808	S cerevisiae apopt
c 398	7	0.5	79	22	AAU54703	Propionibacterium	471	7	0.5	100	23	AAG78368	Human NMDA-R2B C-t
c 399	7	0.5	79	22	AAU56267	Propionibacterium	c 472	7	0.5	101	20	AAU24327	Rabies microti BM
400	7	0.5	79	22	AAU62175	Propionibacterium	c 473	7	0.5	101	21	AAB30218	B. microti BMN1-6
401	7	0.5	80	21	AAG10149	Arabidopsis thalia	474	7	0.5	101	21	AAG10148	Arabidopsis thalia
c 402	7	0.5	80	22	ABG06551	Novel human diagno	c 475	7	0.5	101	23	ABB88963	Babesia microti an
c 403	7	0.5	80	23	ABP79278	Human prostate spe	476	7	0.5	102	22	AAU39200	Propionibacterium
404	7	0.5	81	16	AAU70158	Streptococcus pneu	477	7	0.5	102	22	AAM87171	Human immune/haema
c 405	7	0.5	81	20	AAU76537	Human ovarian tumo	478	7	0.5	102	22	AAO07991	Human polypeptide
c 406	7	0.5	81	22	AAU31912	Novel human secret	479	7	0.5	103	20	AAU39782	Interleukin-18 rec
407	7	0.5	81	22	AAM88234	Human immune/haema	480	7	0.5	103	20	AAU29191	Amino acid sequenc
c 408	7	0.5	81	22	AAM89768	Human immune/haema	c 481	7	0.5	103	20	AAU12429	Human 5' EST secre
c 409	7	0.5	81	22	AAM91747	Human immune/haema	482	7	0.5	103	21	AAG33671	Arabidopsis thalia
410	7	0.5	82	21	AAG00854	Human secreted pro	c 483	7	0.5	105	20	AAU24328	Babesia microti BM
411	7	0.5	83	21	AAG23801	Arabidopsis thalia	c 484	7	0.5	105	20	AAB30219	B. microti BMN1-6
c 412	7	0.5	83	22	AAU52126	Propionibacterium	c 485	7	0.5	105	22	ABP71450	Drosophila melanog
413	7	0.5	84	21	AAG57122	Arabidopsis thalia	c 486	7	0.5	105	23	ABB88964	Babesia microti an
414	7	0.5	86	22	AAM86217	Human immune/haema	c 487	7	0.5	106	20	AAU12739	Human 5' EST secre
c 415	7	0.5	87	20	AAU02893	Fragment of human	488	7	0.5	106	22	AAU63316	Propionibacterium
c 416	7	0.5	87	22	AAU66176	Propionibacterium	489	7	0.5	106	22	ABBI7891	Human nervous syst
417	7	0.5	87	22	AAE01272	Human gene 7 encod	c 490	7	0.5	107	22	AAU59177	Propionibacterium
c 418	7	0.5	87	23	ABG63698	Human albumin fusi	491	7	0.5	107	22	AAU62833	Propionibacterium
c 419	7	0.5	88	21	AAG44944	Zea mays protein f	492	7	0.5	107	22	AAE03402	Human gene 10 enco
c 420	7	0.5	88	22	AAU48338	Propionibacterium	493	7	0.5	107	23	ABG63684	Human albumin fusi
421	7	0.5	88	22	AAU61588	Propionibacterium	c 494	7	0.5	108	22	AAM89714	Human immune/haema
c 422	7	0.5	88	22	AAU61588	Propionibacterium	c 495	7	0.5	109	18	AAW20446	H. pylori transpor
423	7	0.5	88	22	AAG75640	Human colon cancer	c 496	7	0.5	109	18	AAW24662	Novel human diagno
c 424	7	0.5	89	22	ABP69846	Drosophila melanog	c 497	7	0.5	109	22	ABG05912	Streptococcus pneu
c 425	7	0.5	89	22	AAM54691	Propionibacterium	c 498	7	0.5	110	19	AAW55113	Propionibacterium
c 426	7	0.5	89	22	AAM58314	Human brain expres	499	7	0.5	110	22	AAU54045	Human secreted pro
c 427	7	0.5	89	22	AAM70788	Human bone marrow	c 500	7	0.5	110	23	AAB90570	Human albumin fusi
c 428	7	0.5	89	22	AAM31087	Peptide #5124 enco	c 501	7	0.5	110	23	AAG65469	Human secreted pro
c 429	7	0.5	89	22	AAM06195	Peptide #4877 enco	c 502	7	0.5	110	23	ABP54607	S. pneumoniae SP06
c 430	7	0.5	90	21	AAG17387	Arabidopsis thalia	c 503	7	0.5	111	22	ABP54607	B. microti antigen
c 431	7	0.5	90	21	AAG44943	Zea mays protein f	c 504	7	0.5	111	22	AAU41399	Propionibacterium
c 432	7	0.5	90	22	ABBI5742	Human nervous syst	c 505	7	0.5	111	22	ABG13261	Novel human diagno
c 433	7	0.5	91	22	AAO06967	Human polypeptide	c 506	7	0.5	111	22	AAU14804	Novel bone marrow

c 507	111	23	ABP00417	Human ORFX protein	580	7	0.5	131	21	AAG12896	Arabidopsis thalia
c 508	112	22	ABB04681	B. microti antigen	c 581	7	0.5	131	22	ABB11734	Human GPCR homolog
c 509	112	22	AU57069	Propionibacterium	c 582	7	0.5	131	22	AAU18534	Human cytoskeletal
c 510	112	22	AAE03551	Human trefoil doma	c 583	7	0.5	131	22	AAE76585	Corynebacterium gl
c 511	114	21	AAG11367	Arabidopsis thalia	c 584	7	0.5	131	22	AAU21809	Novel human neopla
c 512	115	20	AAV45268	Human secreted pro	585	7	0.5	132	22	AAU03513	Novel human diagno
c 513	115	20	AAW88609	Secreted protein e	586	7	0.5	132	22	AAU01227	Human polypeptide
c 514	115	22	ABE50376	Human secreted pro	587	7	0.5	133	20	AAV30737	Amino acid sequenc
c 515	115	22	ABE17087	Human nervous syst	588	7	0.5	133	22	AAU60129	Propionibacterium
c 516	116	19	AAW54001	Anti-CD4 antibody	c 589	7	0.5	133	22	AAU64797	Propionibacterium
c 517	116	21	AGL16996	Arabidopsis thalia	c 590	7	0.5	134	8	AAV711175	First protein chal
c 518	116	23	AAE18340	Human B7-like prot	c 591	7	0.5	134	9	AAE80018	Sequence of the 18
c 519	117	22	AGG09163	Novel human diagno	592	7	0.5	134	21	AAE56272	Human secreted pro
c 520	117	23	ABG35320	Thrombopoietin ago	c 593	7	0.5	134	22	AAE68139	Amino acid sequenc
c 521	117	23	ABG35320	Human MPL #1. Hom	c 594	7	0.5	134	23	AAE51942	Human TGFbeta prot
c 522	119	22	ABE69173	Drosophila melanog	595	7	0.5	135	21	AAE33445	Human PRO1155 prot
c 523	119	22	ABE69162	Drosophila melanog	596	7	0.5	135	21	AAE66739	Membrane-bound pro
c 524	119	22	AAU51147	Propionibacterium	597	7	0.5	135	22	AAU29245	Human PRO polypept
c 525	119	22	AGG89199	Human secreted pro	598	7	0.5	135	22	AAO01506	Human polypeptide
c 526	120	13	AAE27497	WN1 222-5 antibody	c 599	7	0.5	135	22	AGG92481	C glutamicum prote
c 527	120	21	AAW94275	Human eIF-4E-bind	600	7	0.5	135	22	AAE65262	Human PRO1155 (UNQ
c 528	120	21	AAW96148	Human eIF-4E bind	601	7	0.5	135	23	AAE95507	Human angio genesis
c 529	120	22	AAU44645	Propionibacterium	602	7	0.5	135	23	ABE84901	Human PRO1155 prot
c 530	120	22	ABE03155	Human musculoskele	603	7	0.5	135	23	AAU83645	Human angio genesis
c 531	121	19	AAW75228	Human secreted pro	c 604	7	0.5	136	21	AAE93275	Amino acid sequenc
c 532	121	19	AAW75212	Human secreted pro	605	7	0.5	136	22	ABG07959	Novel human diagno
c 533	121	20	AAW97213	A human neurokln p	c 606	7	0.5	136	22	ABG12201	Novel human diagno
c 534	121	20	AAW74413	HPMB091 protein se	c 607	7	0.5	136	23	ABG35323	Thrombopoietin ago
c 535	121	22	ABG10328	Novel human diagno	c 608	7	0.5	137	19	AAW77671	30S ribosomal prot
c 536	121	22	ABG82176	S. epidermidis ope	609	7	0.5	137	21	AAE41833	Arabidopsis thalia
c 537	122	22	AAE82380	Human neurokinin B	c 610	7	0.5	137	21	AAV71471	Human prostaglandi
c 538	122	20	AAW96144	Human preprotachyk	c 611	7	0.5	137	22	ABE11878	Human neurokinin B
c 539	122	21	AGG07128	Arabidopsis thalia	612	7	0.5	139	17	AAE89134	Human early placen
c 540	122	21	AAE58442	Arabidopsis thalia	613	7	0.5	139	18	AAW17676	Human relaxin-rela
c 541	122	21	AGG59389	Arabidopsis thalia	614	7	0.5	139	19	AAW69168	Zinsi protein. Ho
c 542	122	21	AGG61242	Arabidopsis thalia	615	7	0.5	139	20	AAE26926	Human insulin-like
c 543	122	21	AGB73201	Inhibin balpha C-t	616	7	0.5	139	20	AAW99574	Human early placen
c 544	123	23	ABP38750	Staphylococcus epi	c 617	7	0.5	139	22	ABG05772	Novel human diagno
c 545	123	20	AAV39787	Interleukin-18 rec	c 618	7	0.5	140	22	AAO04333	Human polypeptide
c 546	123	21	AAE53457	Human colon cancer	c 619	7	0.5	140	22	AAE44909	zea mays protein f
c 547	123	21	AAE333846	Human secreted pro	c 620	7	0.5	140	22	AAE82356	S. epidermidis ope
c 548	123	21	AAE40493	Human ORFX ORF257	c 621	7	0.5	140	23	AAW47842	Human protein sequ
c 549	123	21	AAE12897	Arabidopsis thalia	c 622	7	0.5	141	20	AAE27815	Human secreted pro
c 550	123	22	AAE90658	Human lipocolon (L	c 623	7	0.5	141	22	AAE64529	Propionibacterium
c 551	123	23	ABP01509	Human ORFX protein	c 624	7	0.5	141	22	AAO02557	Human polypeptide
c 552	124	21	AGG07127	Arabidopsis thalia	c 625	7	0.5	143	22	AAE68333	Human lipocalin ho
c 553	124	21	AGG57121	Arabidopsis thalia	c 626	7	0.5	144	23	ABF39916	Staphylococcus epi
c 554	124	21	AGG58441	Arabidopsis thalia	c 627	7	0.5	145	21	AAE12895	Arabidopsis thalia
c 555	124	21	AGG59295	Arabidopsis thalia	c 628	7	0.5	145	21	AAE21533	Arabidopsis thalia
c 556	124	21	AGG59388	Arabidopsis thalia	c 629	7	0.5	145	21	AAE25285	Arabidopsis thalia
c 557	124	21	AGG61241	Arabidopsis thalia	c 630	7	0.5	145	21	AAE29658	Arabidopsis thalia
c 558	124	22	AAE64677	Human secreted pro	c 631	7	0.5	145	22	ABE69035	Drosophila melanog
c 559	125	21	AAE34558	Arabidopsis thalia	c 632	7	0.5	146	21	AAE24496	Arabidopsis thalia
c 560	126	21	AAE44910	zea mays protein f	c 633	7	0.5	146	21	AAE36364	Arabidopsis thalia
c 561	126	22	AAU50028	Propionibacterium	634	7	0.5	146	21	AAE43677	Arabidopsis thalia
c 562	126	22	AAE71848	Human olfactory re	635	7	0.5	146	21	AAE59299	Arabidopsis thalia
c 563	127	20	AAV24088	Salmonella typhimu	636	7	0.5	146	22	ABG15019	Novel human diagno
c 564	127	20	AAU55399	Propionibacterium	637	7	0.5	146	22	ABG27696	Novel human diagno
c 565	127	22	ABG07966	Novel human diagno	638	7	0.5	146	22	AAE87060	Human immune/haema
c 566	128	21	AAE19639	Mouse monoclonal a	639	7	0.5	146	22	AAU16923	Human novel secret
c 567	128	21	AAE19643	Chimeric antibody	640	7	0.5	147	22	ABE52739	Escherichia coli p
c 568	128	21	AAE30201	B. microti BMNI-16	641	7	0.5	148	21	AAE23800	Arabidopsis thalia
c 569	128	22	ABE68852	Drosophila melanog	642	7	0.5	148	22	AAO06229	Human polypeptide
c 570	128	22	AAU51285	Propionibacterium	643	7	0.5	149	21	AAE18616	Amino acid sequenc
c 571	128	22	AGG01132	Novel human diagno	c 644	7	0.5	149	21	AAE36363	Arabidopsis thalia
c 572	128	23	ABE88947	Babesia microti an	c 645	7	0.5	149	22	AAE00855	Human bone marrow
c 573	129	20	AAV37136	Amino acid sequenc	646	7	0.5	149	22	AAE82675	S. epidermidis ope
c 574	129	21	AAE13005	Arabidopsis thalia	647	7	0.5	150	21	AAE24495	Arabidopsis thalia
c 575	129	21	AAE34557	Arabidopsis thalia	648	7	0.5	150	21	AAE43676	Arabidopsis thalia
c 576	129	21	AGG00096	Human secreted pro	c 649	7	0.5	150	22	ABG01944	Novel human diagno
c 577	129	22	ABG10326	Novel human diagno	c 650	7	0.5	151	21	AAE42294	Human ORFX ORF2058
c 578	130	22	ABG06878	Novel human diagno	c 651	7	0.5	151	21	AAE18952	zea mays protein f
c 579	130	23	ABE55608	Lactococcus lactis	652	7	0.5	151	23	ABP01033	Human ORFX protein

c 653	7	0.5	151	23	ABB54403	Lactococcus lactis	726	7	0.5	173	22	AAU38852	Novel subtilisin h
c 654	7	0.5	152	21	AAQ24579	Arabidopsis thalia	727	7	0.5	173	22	AAU38853	Novel subtilisin h
c 655	7	0.5	152	21	AAQ32508	Arabidopsis thalia	728	7	0.5	173	22	AAU38856	Novel subtilisin h
c 656	7	0.5	152	22	AAU29597	Novel human secret	c 729	7	0.5	174	20	AAV39208	M. tuberculosis an
c 657	7	0.5	153	22	ABB64614	Drosophila melanog	c 730	7	0.5	174	20	AAV39065	M. tuberculosis re
c 658	7	0.5	153	22	ABB71795	Drosophila melanog	731	7	0.5	174	22	AAU46916	Propionibacterium
c 659	7	0.5	154	20	AAW30642	A. thaliana xylan	732	7	0.5	174	22	AAO08955	Human pancreatic c
c 660	7	0.5	154	22	AAW30642	Human secreted pro	733	7	0.5	175	21	AAW54279	Human pancreatic c
c 661	7	0.5	155	21	AAW18906	Amino acid sequenc	c 734	7	0.5	176	22	AAO02230	Human polypeptide
c 662	7	0.5	155	21	AAW18906	Human cancer assoc	c 735	7	0.5	177	20	AAU34610	Chlamydia pneumoni
c 663	7	0.5	156	23	ABF40578	Staphylococcus epi	c 736	7	0.5	177	23	ABB90534	Chlamydia pneumoni
c 664	7	0.5	156	23	ABB90529	Human polypeptide	c 737	7	0.5	178	20	AAV41229	M. tuberculosis yv
c 665	7	0.5	157	21	AAW86238	Human secreted pro	738	7	0.5	178	21	AAU09108	Arabidopsis thalia
c 666	7	0.5	157	22	ABB68297	Drosophila melanog	739	7	0.5	178	21	AAU10776	Arabidopsis thalia
c 667	7	0.5	158	21	AAW24690	Plant SDF encoded	740	7	0.5	178	22	ABG11517	Novel human diago
c 668	7	0.5	158	21	AAW25075	Plant SDF encoded	741	7	0.5	178	22	AAW85803	Human pPARGamma-PA
c 669	7	0.5	158	21	AAW25075	Arabidopsis thalia	742	7	0.5	179	21	AAV94465	Human polypeptide
c 670	7	0.5	158	21	AAW25075	Arabidopsis thalia	c 743	7	0.5	179	22	AAO01805	Human polypeptide
c 671	7	0.5	158	22	ABG01676	Novel human diago	c 744	7	0.5	180	21	AAW36359	Human cancer assoc
c 672	7	0.5	159	21	AAW18906	Arabidopsis thalia	c 745	7	0.5	180	23	ABB89716	Human polypeptide
c 673	7	0.5	159	23	ABB89623	Human polypeptide	c 746	7	0.5	180	23	ABW54709	Lactococcus lactis
c 674	7	0.5	161	22	AAW65024	Human secreted pro	c 747	7	0.5	182	20	AAW60131	Human endometrium
c 675	7	0.5	162	21	AAW07126	Arabidopsis thalia	c 748	7	0.5	182	22	AAU67125	Novel central nerv
c 676	7	0.5	162	21	AAW07126	Arabidopsis thalia	749	7	0.5	182	22	ABB67070	Drosophila melanog
c 677	7	0.5	162	21	AAW07126	Arabidopsis thalia	c 750	7	0.5	182	22	AAU18538	Human cytoskeletal
c 678	7	0.5	162	22	ABG28982	Novel human diago	c 751	7	0.5	182	22	AAU21722	Novel human neopla
c 679	7	0.5	162	22	ABG28982	Novel human diago	c 752	7	0.5	184	22	AAU60335	Propionibacterium
c 680	7	0.5	162	22	AAU18632	Human lung antigen	c 753	7	0.5	184	22	AAU12290	Human PRO6093 poly
c 681	7	0.5	163	21	AAW75681	Neisseria gonorrhoe	c 754	7	0.5	184	23	AAW68321	Human lipocalin ho
c 682	7	0.5	163	22	AAU49350	Propionibacterium	c 755	7	0.5	184	23	AAW69554	Human G protein-co
c 683	7	0.5	163	22	AAU59542	Propionibacterium	c 756	7	0.5	185	21	AAW07829	Arabidopsis thalia
c 684	7	0.5	164	21	AAW32939	Pinus radiata tran	c 757	7	0.5	185	21	AAW27605	Arabidopsis thalia
c 685	7	0.5	164	22	ABW52672	Escherichia coli p	c 758	7	0.5	185	23	ABP40635	Staphylococcus epi
c 686	7	0.5	165	22	ABW52672	Novel human diago	c 759	7	0.5	185	23	ABP30037	Streptococcus poly
c 687	7	0.5	166	21	ABG24494	Arabidopsis thalia	c 760	7	0.5	185	23	ABW53953	Lactococcus lactis
c 688	7	0.5	166	22	ABG13630	Novel human diago	c 761	7	0.5	186	22	ABG01371	Novel human diago
c 689	7	0.5	166	23	ABP28242	Streptococcus poly	c 762	7	0.5	186	22	AAW94294	Human protein sequ
c 690	7	0.5	169	21	AAU10670	Arabidopsis thalia	c 763	7	0.5	187	21	AAW36362	Arabidopsis thalia
c 691	7	0.5	169	21	AAW24578	Arabidopsis thalia	c 764	7	0.5	187	21	ABG19655	Novel human diago
c 692	7	0.5	169	21	AAW24578	Arabidopsis thalia	c 765	7	0.5	187	22	ABG19656	Novel human diago
c 693	7	0.5	169	21	AAW24578	Arabidopsis thalia	c 766	7	0.5	187	23	AAO22163	Ramoplanin biosync
c 694	7	0.5	170	21	AAW96467	Partial Veronia 4 -	c 767	7	0.5	187	23	ABW92305	Herbicidially activ
c 695	7	0.5	171	22	AAU47971	Propionibacterium	c 768	7	0.5	188	21	AAW31499	Arabidopsis thalia
c 696	7	0.5	171	22	AAW28269	Human peptide #920	c 769	7	0.5	188	22	ABW59566	Drosophila melanog
c 697	7	0.5	171	22	ABW29093	Peptide #1744 enco	c 770	7	0.5	188	22	AAW66790	Mouse ZALPHA29 pro
c 698	7	0.5	171	22	ABW30969	Peptide #3620 enco	c 771	7	0.5	190	20	AAW88708	Secreted protein e
c 699	7	0.5	171	22	ABW33448	Peptide #954 enco	c 772	7	0.5	190	21	AAW46323	Arabidopsis thalia
c 700	7	0.5	171	22	ABW34254	Peptide #1760 enco	c 773	7	0.5	190	22	ABW63097	Drosophila melanog
c 701	7	0.5	171	22	ABW36162	Peptide #3668 enco	c 774	7	0.5	190	22	ABW50475	Human secreted pro
c 702	7	0.5	171	22	ABW18904	Protein #903 enco	c 775	7	0.5	190	22	AAW78656	Human protein SEQ
c 703	7	0.5	171	22	ABW19690	Protein #1689 enco	c 776	7	0.5	190	22	AAU12250	Human PRO4371 poly
c 704	7	0.5	171	22	ABW21546	Protein #3545 enco	c 777	7	0.5	190	22	AAW66789	Human ZALPHA29 pro
c 705	7	0.5	171	22	AAW55047	Human brain expres	c 778	7	0.5	190	23	ABW95572	Human angiotogenesis
c 706	7	0.5	171	22	AAW56944	Human brain expres	c 779	7	0.5	190	23	ABW89971	Human polypeptide
c 707	7	0.5	171	22	AAW67437	Human bone marrow	c 780	7	0.5	190	23	ABW84966	Human PRO4371 prot
c 708	7	0.5	171	22	AAW69331	Human bone marrow	c 781	7	0.5	191	20	AAW73377	Human HPDDV78 prot
c 709	7	0.5	171	22	AAW15263	Peptide #1697 enco	c 782	7	0.5	191	21	AAW54698	Arabidopsis thalia
c 710	7	0.5	171	22	AAW17160	Peptide #3594 enco	c 783	7	0.5	192	22	AAU57378	Propionibacterium
c 711	7	0.5	171	22	AAW26908	Peptide #945 enco	c 784	7	0.5	192	22	AAW72321	Human olfactory re
c 712	7	0.5	171	22	AAW27728	Peptide #1765 enco	c 785	7	0.5	192	23	AAO14525	Novel human EBI-3-
c 713	7	0.5	171	22	AAW29651	Peptide #3688 enco	c 786	7	0.5	192	23	AAW68325	Human lipocalin ho
c 714	7	0.5	171	22	AAW30008	Peptide #1690 enco	c 787	7	0.5	193	13	AAW25663	Recombinant porcin
c 715	7	0.5	171	22	AAW04854	Peptide #3536 enco	c 788	7	0.5	193	21	AAW32994	Pinus radiata tran
c 716	7	0.5	171	23	ABG37060	Human peptide enco	c 789	7	0.5	193	21	AAW74590	Neisseria meningit
c 717	7	0.5	171	23	ABG38947	Human peptide enco	c 790	7	0.5	193	22	AAU40711	Propionibacterium
c 718	7	0.5	172	21	AAW30332	Arabidopsis thalia	c 791	7	0.5	193	22	ABG22169	Novel human diago
c 719	7	0.5	172	22	AAW69507	Human purified sec	c 792	7	0.5	194	20	AAW07465	Mouse TSI0q23.3 ge
c 720	7	0.5	173	20	AAW86322	Kidney injury asso	c 793	7	0.5	194	22	ABW66978	Drosophila melanog
c 721	7	0.5	173	22	AAW38795	Novel subtilisin h	c 794	7	0.5	194	22	ABG12771	Novel human diago
c 722	7	0.5	173	22	AAW38795	Novel subtilisin h	c 795	7	0.5	194	23	ABP42183	Human ovarian anti
c 723	7	0.5	173	22	AAW38811	Novel subtilisin h	c 796	7	0.5	195	21	AAW70150	Staphylococcus aur
c 724	7	0.5	173	22	AAW38839	Novel subtilisin h	c 797	7	0.5	195	22	ABW60084	Drosophila melanog
c 725	7	0.5	173	22	AAW38849	Novel subtilisin h	c 798	7	0.5	195	22	ABW12318	Human secreted pro

c 799	7	0.5	195	22	AAW79640	Human protein SEQ	c 872	7	0.5	222	21	AAAG07828	Arabidopsis thalia
c 800	7	0.5	195	22	ABP95863	Human protein sequ	c 873	7	0.5	222	21	AAAG16141	Arabidopsis thalia
c 801	7	0.5	195	22	ABP60776	Gene 47 related be	c 874	7	0.5	222	21	AAAG27604	Arabidopsis thalia
c 802	7	0.5	195	23	ABP06799	Human ngPCR-Seq101	c 875	7	0.5	222	22	AAU31028	Novel human secret
c 803	7	0.5	196	22	AAU87444	Novel central nerv	c 876	7	0.5	222	22	AAU03629	Group B Streptococ
c 804	7	0.5	196	22	AAU35232	Enterococcus faeca	c 877	7	0.5	222	23	ABP26468	Streptococcus poly
c 805	7	0.5	196	22	AAU18648	Renal and cardiova	c 878	7	0.5	223	21	AAU57038	Human prostate can
c 806	7	0.5	197	22	ABG08070	Novel human diagno	c 879	7	0.5	223	21	AAAG24725	Arabidopsis thalia
c 807	7	0.5	198	22	AAU63455	Propionibacterium	c 880	7	0.5	223	22	ABG23067	Novel human diagno
c 808	7	0.5	198	22	AAU63455	Propionibacterium	c 881	7	0.5	223	22	AAU40614	Human polypeptide
c 809	7	0.5	198	22	AAU69495	Human purified sec	c 882	7	0.5	224	20	AAU74086	Human prostate tum
c 810	7	0.5	198	22	AAU17112	Novel signal trans	c 883	7	0.5	224	22	ABG01822	Novel human diagno
c 811	7	0.5	198	23	AAU68329	Human lipocalin ho	c 884	7	0.5	225	22	ABG15290	Gene 3 human secre
c 812	7	0.5	200	20	AAU88634	Secreted protein e	c 885	7	0.5	225	13	ABP42795	Human ovarian anti
c 813	7	0.5	200	21	AAU53277	Human colon cancer	c 886	7	0.5	226	12	AAU13757	Prolactin. Mus mu
c 814	7	0.5	200	22	AAU33308	Staphylococcus aur	c 887	7	0.5	226	20	AAU39788	Interleukin-18 rec
c 815	7	0.5	200	22	AAU36985	Staphylococcus aur	c 888	7	0.5	226	21	AAU19331	Arabidopsis thalia
c 816	7	0.5	200	22	AAU37185	Staphylococcus aur	c 889	7	0.5	226	22	ABG03260	Novel human diagno
c 817	7	0.5	200	22	ABP50401	Human secreted pro	c 890	7	0.5	226	23	ABP06325	Human ORFX protein
c 818	7	0.5	200	23	AAO20048	Staphylococcus aur	c 891	7	0.5	227	20	AAU05312	S. aureus protein
c 819	7	0.5	200	23	AAO20069	Staphylococcus aur	c 892	7	0.5	227	22	ABG71814	Drosophila melanog
c 820	7	0.5	202	21	AAU96806	Streptococcus agal	c 893	7	0.5	227	22	ABG04921	Novel human diagno
c 821	7	0.5	202	22	AAU63393	Propionibacterium	c 894	7	0.5	228	23	ABP33032	Human ORF2005 prot
c 822	7	0.5	202	22	AAU66462	Propionibacterium	c 895	7	0.5	228	20	ABP55417	Lactococcus lactis
c 823	7	0.5	203	22	ABG12756	Novel human diagno	c 896	7	0.5	229	23	AAU27341	Group B Streptococ
c 824	7	0.5	204	21	AAU24726	Arabidopsis thalia	c 897	7	0.5	229	21	AAU31498	Arabidopsis thalia
c 825	7	0.5	204	23	ABP77327	Human growth hormo	c 898	7	0.5	230	22	AAU81587	S. epidermidis ope
c 826	7	0.5	205	21	AAU09892	Human glycine tran	c 899	7	0.5	231	12	AAU13518	M.ivanovii CORA.
c 827	7	0.5	205	21	AAU21164	Human glycine tran	c 900	7	0.5	231	21	AAU26310	Arabidopsis thalia
c 828	7	0.5	205	21	AAU21165	Human glycine tran	c 901	7	0.5	231	21	AAU37677	Arabidopsis thalia
c 829	7	0.5	205	21	AAU21166	Human glycine tran	c 902	7	0.5	231	21	AAU93380	Amino acid sequenc
c 830	7	0.5	205	21	AAU21167	Human glycine tran	c 903	7	0.5	231	21	AAU93300	Amino acid sequenc
c 831	7	0.5	205	21	AAU21168	Human glycine tran	c 904	7	0.5	231	21	AAU74392	Neisseria gonorrhoe
c 832	7	0.5	205	21	AAU21169	Human glycine tran	c 905	7	0.5	231	21	AAU74393	Neisseria meningit
c 833	7	0.5	205	22	ABP73956	Canine mutant RPE6	c 906	7	0.5	231	21	AAU74394	Neisseria meningit
c 834	7	0.5	206	21	AAU21119	Hydrophobic domain	c 907	7	0.5	231	22	AAU58594	Propionibacterium
c 835	7	0.5	206	22	ABP71052	Drosophila melanog	c 908	7	0.5	231	23	AAU20325	Human B7-H8 protei
c 836	7	0.5	207	20	AAU39784	Interleukin-18 rec	c 909	7	0.5	232	19	AAU75021	Human secreted pro
c 837	7	0.5	207	22	ABG29046	Novel human diagno	c 910	7	0.5	232	19	AAU75021	Human secreted pro
c 838	7	0.5	208	21	AAU16142	Arabidopsis thalia	c 911	7	0.5	233	21	AAU22269	Zea mays protein f
c 839	7	0.5	208	21	AAU26312	Arabidopsis thalia	c 912	7	0.5	233	21	AAU09752	IGFBP-1 amino acid
c 840	7	0.5	208	21	AAU37679	Arabidopsis thalia	c 913	7	0.5	234	15	AAU62046	Human mature IGFBP
c 841	7	0.5	208	22	AAU18536	Human cytoskeletal	c 914	7	0.5	234	21	AAU09617	Insulin like growt
c 842	7	0.5	208	22	AAU23725	Human EST encoded	c 915	7	0.5	235	21	AAU41562	Arabidopsis thalia
c 843	7	0.5	208	22	AAU23729	Human EST encoded	c 916	7	0.5	236	21	AAU30331	Arabidopsis thalia
c 844	7	0.5	208	22	AAU21823	Novel human neopla	c 917	7	0.5	237	21	AAU57100	Human prostate can
c 845	7	0.5	208	23	ABP30274	Streptococcus poly	c 918	7	0.5	237	21	AAU07827	Arabidopsis thalia
c 846	7	0.5	209	21	AAU42974	Human ORFX ORF2738	c 919	7	0.5	237	21	AAU27603	Arabidopsis thalia
c 847	7	0.5	210	22	ABP70342	Drosophila melanog	c 920	7	0.5	237	21	AAU36022	Arabidopsis thalia
c 848	7	0.5	211	21	AAU19332	Arabidopsis thalia	c 921	7	0.5	238	20	AAU07464	Zea mays protein f
c 849	7	0.5	211	22	AAU27475	Human G-Protein Co	c 922	7	0.5	238	20	AAU07022	Human TS10q23.3 ge
c 850	7	0.5	213	23	AAU97619	D. Immittis cystein	c 923	7	0.5	238	22	ABG05347	Breast cancer asso
c 851	7	0.5	215	23	ABG64839	Human albumin fusi	c 924	7	0.5	238	23	AAU50339	Novel human diagno
c 852	7	0.5	215	23	AAU96215	Human secreted pro	c 925	7	0.5	241	20	AAU06715	ETEC CS4 pilus Csa
c 853	7	0.5	215	23	AAU96235	Human secreted pro	c 926	7	0.5	241	23	AAU73982	Human prostate tum
c 854	7	0.5	216	23	ABP25952	Streptococcus poly	c 927	7	0.5	242	23	ABP40309	Cysteine protease
c 855	7	0.5	217	22	AAU75021	Human colon cancer	c 928	7	0.5	243	23	ABP47788	Staphylococcus epi
c 856	7	0.5	218	19	AAU70322	Secreted protein C	c 929	7	0.5	244	20	AAU06716	Listeria monocytog
c 857	7	0.5	219	22	ABU12045	Human channel-rela	c 930	7	0.5	245	20	AAU06715	Antibody 12B5 sing
c 858	7	0.5	219	22	AAU40605	Human polypeptide	c 931	7	0.5	245	22	AAU25480	Antibody 10D10 sin
c 859	7	0.5	219	23	ABG64782	Human albumin fusi	c 932	7	0.5	245	22	AAU25480	Human protein sequ
c 860	7	0.5	219	23	AAU17197	Human secreted pro	c 933	7	0.5	245	22	AAU92570	C glutamic prote
c 861	7	0.5	219	23	AAU17230	Human secreted pro	c 934	7	0.5	246	21	AAU76520	Corynebacterium gl
c 862	7	0.5	220	21	AAU26311	Arabidopsis thalia	c 935	7	0.5	247	21	AAU15540	Human immune syste
c 863	7	0.5	220	21	AAU37678	Arabidopsis thalia	c 936	7	0.5	247	21	AAU31658	Arabidopsis thalia
c 864	7	0.5	220	22	AAU02481	Arabidopsis thalia	c 937	7	0.5	249	18	AAU31220	HIV-TAT protein tr
c 865	7	0.5	220	22	AAU65001	Human secreted pro	c 938	7	0.5	249	19	AAU26460	Protein of the inv
c 866	7	0.5	220	23	ABP40228	Staphylococcus epi	c 939	7	0.5	249	19	AAU76166	Seg ID 57 from US5
c 867	7	0.5	220	23	AAU78292	Human Pregnane X R	c 940	7	0.5	251	21	AAU16369	Protein #2 used fo
c 868	7	0.5	221	21	AAU18928	A novel polypeptid	c 941	7	0.5	251	21	AAU45627	Arabidopsis thalia
c 869	7	0.5	221	22	AAU29267	Human PRO polypept	c 942	7	0.5	251	22	AAU34822	E. coli cellular p
c 870	7	0.5	221	22	AAU47451	SMAP-1. Homo sapi	c 943	7	0.5	251	22	AAU06774	Human polypeptide
c 871	7	0.5	221	22	AAU38828	Human polypeptide	c 944	7	0.5	251	23	AAU91449	Haemophilus influe

```

c 945 7 0.5 252 22 ABG17943
c 946 7 0.5 253 21 AAG24724
c 947 7 0.5 253 21 AAG31497
c 948 7 0.5 253 22 AAM78946
c 949 7 0.5 253 22 AAM77995
c 950 7 0.5 253 22 AAU01525
c 951 7 0.5 253 23 AB47442
c 952 7 0.5 254 20 AAX29166
c 953 7 0.5 254 22 AAU45815
c 954 7 0.5 254 22 AAM39922
c 955 7 0.5 254 22 AAB94528
c 956 7 0.5 254 22 AAB61163
c 957 7 0.5 255 22 AAM79497
c 958 7 0.5 255 22 AAG89821
c 959 7 0.5 255 23 AAG60125
c 960 7 0.5 255 23 AAB20324
c 961 7 0.5 256 22 AAU41531
c 962 7 0.5 256 22 AAB30807
c 963 7 0.5 257 22 AAG04655
c 964 7 0.5 258 10 AAP93154
c 965 7 0.5 258 21 AAG17221
c 966 7 0.5 258 21 AAG21532
c 967 7 0.5 258 21 AAG25284
c 968 7 0.5 258 21 AAG29657
c 969 7 0.5 258 21 AAG46911
c 970 7 0.5 258 23 ABB91111
c 971 7 0.5 258 23 AAE18337
c 972 7 0.5 259 10 AAP91868
c 973 7 0.5 259 21 AAG32408
c 974 7 0.5 259 22 AAM25547
c 975 7 0.5 259 22 AAM41708
c 976 7 0.5 260 11 AAR08332
c 977 7 0.5 260 20 AAV35370
c 978 7 0.5 260 21 AAG12858
c 979 7 0.5 261 22 AAU40132
c 980 7 0.5 261 22 AAG15226
c 981 7 0.5 261 22 AAU30474
c 982 7 0.5 261 22 AAB67762
c 983 7 0.5 263 19 AAW56277
c 984 7 0.5 263 20 AAV24339
c 985 7 0.5 263 21 AAB30188
c 986 7 0.5 263 21 AAG41361
c 987 7 0.5 263 22 AAG25026
c 988 7 0.5 263 22 AAG92021
c 989 7 0.5 263 23 ABB76518
c 990 7 0.5 264 21 AAG40100
c 991 7 0.5 265 22 AAB80417
c 992 7 0.5 266 23 ABB35329
c 993 7 0.5 267 22 AAB65084
c 994 7 0.5 268 21 AAG06165
c 995 7 0.5 268 21 AAG10669
c 996 7 0.5 269 22 AAU22509
c 997 7 0.5 269 22 AAM92464
c 998 7 0.5 270 21 AAB11911
c 999 7 0.5 270 21 AAB11930
1000 7 0.5 270 22 ABB28985

```

ALIGNMENTS

```

RESULT 1
AAB92617
ID AAB92617 standard; Protein; 507 AA.
AC AAB92617;

```

```

XX 26-JUN-2001 (first entry)

```

```

XX Human protein sequence SEQ ID NO:10896.

```

```

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

```

```

KW Homo sapiens.

```

```

OS

```

```

XX EPI074617-A2.
PN 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 10896; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX SQ Sequence 507 AA;

```

```

Alignment Scores:
Pred. No.: 0 Length: 507
Score: 494.00 Matches: 494
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.54% Indels: 0
DB: 22 Gaps: 0

```

```

US-09-818-990B-1 (1-3963) x AAB92617 (1-507)

```

```

Qy 1 ATGCAAGACGACACATAGAGCTTCTACATATCTCCATATCTCAGCTTCTAAGAGAGAGCTAT 60
Db 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20

```

```

Qy 61 TTAGCTGAACACGACATCGGGAAACAATGAGAGAGTGCAGCGGAGCCCTCTCCCAAC 120
Db 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40

```

```

Qy 121 CCTTGCCATTTCGGCACTCTCTTGTGGGGCCGCTGAAGAGCGGAGGCCAAGATGACCTT 180
Db 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu 60

```

```
QY 181 CCAGATCTTTCCGCTTTCTGAGCCCAAGAGAATTTAGACGAAAGTGTCAATTTGGCAAGA 240
Db 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
QY 241 CTGGCCATCAATTTACGACCTTTGGAGAGAGCGAGATGAACCTCAAGCTAGAAACGACTT 300
Db 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
QY 301 TCTCCTGATCAGATGAACACTCACCTCAATTTAAGTTTTCAGCTCACTTCTGCCAGGAT 360
Db 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
QY 361 AACCTCTCGAAGTCCCAACCTCTAAAGAAAGCCCGCAGGAGCAAAAGGCCACAGTAT 420
Db 121 AsnProArgSerProThrSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
QY 421 TGTCTGTAACCCAGTCCCAAAAAGTATTTTAAATGAAGCTCCGACTTCATTGAAGAG 480
Db 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
QY 481 CTATCTCTCCCTTTTCAATCCCAACAGCTCCAAAGAGTTAGACCTCGTGCCTGCCAAAC 540
Db 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
QY 541 CACAAGAGTAACCTGGAATCTCAAAACAAAGTTATGCAGGAAACAGCTCCAGTTTCTCA 600
Db 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
QY 601 GATCTGTCAAGAAAGACGAGAAGATCTTCGTTCCTCCATCCCTATCCCTCGCGGATACCAGG 660
Db 201 AspLeuSerGluArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
QY 661 GATAATGAAGTGAATCACGCCCTGGACACAGCAGCAAGCAAGCGGTGAAGCGGACGAG 720
Db 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
QY 721 GCTGCCAGTGCAGCGGTGGTGAGACACTACACACAGGCTTCCTCCCTTCATCTCTGTAC 780
Db 241 AlaAlaSerGluAlaAlaGlyLysArgThrThrProGlySerSerProSerSerLeuTyr 260
QY 781 TATGAGAACCTCTGGGGCAACCTCCCGGTTTCACCAAAAGTTACGGACGACAGAGATT 840
Db 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
QY 841 CCAGAGGAAGTCCAGTACAGTTGGATTGCATAGTGTAGCAATTCACACCTCAAGTA 900
Db 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
QY 901 AGGTGTACTGTGAAGCAAGAGCTTGAATAATTCGCCAGATATTCACATCGTCCAGGCA 960
Db 301 ArgTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
QY 961 GGAATCTGCACCTACTGACCATTTGGGAGAGCTTGAAGAGGACGACGACGCTATTC 1020
Db 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
QY 1021 TGTCTTGTCTTAACATCTATGGACAGATTTCGACTTCTGCTGAGATTTATATAGAGGG 1080
Db 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
QY 1081 GTTCTTCTTCTGACTCAGAGCGACCCCTAACAAAGAGAGATGAATCCAAATCCAGAAG 1140
Db 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
QY 1141 CCAATAGAGTGTATCTCTCCACCTACTCTGCACTCAATCTCCAGCAGTACCCCAA 1200
Db 381 ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln 400
QY 1201 GCCCAGCATTTGTGGCCCAACCTCTGTGGCAACCATCCAGCAGTGTACAGCCCCACC 1260
Db 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnCysGlnSerProThr
```

```
QY 1261 AATTACTTCGAGGATTGGATGGAACACCTATCATTCAGCTCTGTGTTTACAAAGATG 1320
Db 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMet 440
QY 1321 CTCAAAATTTGTCAGCTTCTGAGGGTCAGCTGGTTCCTTTGAATCCAGAGTAAAGGA 1380
Db 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
QY 1381 GCTCCATCTCTAAGTTGAGTGTATAGAGAAGGAGCTTTAATAGAGATTTCTCCAGAT 1440
Db 461 AlaProSerProLysValGluTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
QY 1441 TTTAGGATTTTACAGAAAAACCTCGATCCATCGCAGAGCCA 1482
Db 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
RESULT 2
ABB10330
ID ABB10330 standard; Protein; 264 AA.
XX
AC ABB10330;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 638.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179055P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216890P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225288P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
```

PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 03-SEP-2000; 2000US-229509P.
PR 03-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235848P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.

PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 11-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 03-JAN-2001; 2001US-259678P.
XX
XX

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-476161/51.

DR N-PSDB; ABA06552.

XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -

XX Claim 11; SEQ ID NO: 638; 859pp + Sequence Listing; English.

XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.

XX Sequence 264 AA;

Alignment Scores:
Pred. No.: 9,49e-232 Length: 264
Score: 261.00 Matches: 261
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.83% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x ABB10330 (1-264)

QY 3178 GAAAGAGACAAAGAGCCCTACAGGAACGCTTTTTCGACACACATTTCTCGAGGCTCCT 3237

Db 4 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 23

QY 3238 GGGATATGCTAGCTCATGAGGGCGCCTCTGTGCGGTGGAGCTGTAAAGTGAGTGGTTTA 3297

Db 24 GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeu 43

QY 3298 CCGCCCCCGAGCTGACATGGCTACTCAATGGCCAACTGTGTACCATGCTCCAC 3357

Db 44 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 63

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
XX
XX
XX

DR WPI; 2001-457725/49.
DR N-PSDB; AAS28777.
XX Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX Claim 11; SEQ ID No 134; 551pp; English.
XX Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
CC invention. The polypeptides and their associated polynucleotides can be
CC used to diagnose a pathological condition or a susceptibility to a
CC pathological condition in a subject by determining the presence or
CC absence of a mutation in a DNA sequence or determining the presence or
CC amount of expression of the protein. Alternatively the identification of
CC a binding partner to a sequence allows determination of changes in
CC protein activity. The sequences can be used as research tools for
CC receptors or other signal transduction pathway proteins that interact
CC with the polypeptides of the invention and can be used to treat, prevent
CC or diagnose various types of disorders such as neurological disorders,
CC cardiovascular disorders, gastrointestinal disorders, reproductive
CC disorders, immune system disorders, renal disorders, muscular disorders,
CC pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 261 AA;

Alignment Scores:
Pred. No.: 4.73e-156 Length: 261
Score: 179.00 Matches: 179
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.60% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x AAU17989 (1-261)

QY 3178 GAAAGAGACAAAGAGCCCTACAGGACGCTTTTCGACACATTTCTCGAGGCTCCT 3237
Db 1 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 20

QY 3238 GGGATATGCTAGCTCATGAGGGCGCTCTGTCGGCTGGACTGTAAGGTGAGTGTTTA 3297
Db 21 GlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeu 40

QY 3298 CGGCCCCGGAGTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCAC 3357
Db 41 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 60

QY 3358 AAGATGCTGTCAGGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCAGCGC 3417
Db 61 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuLeuLeuLeuGlnArg 80

QY 3418 GACGCGAGGACCTATAAGTGCATCGCTACCAACAAACCGGCGAGAATCTTTTAGTCTG 3477
Db 81 AspAlaGlyThrTyrlsCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 100

QY 3478 GAGCTCTCTAGTAGCCAAAGAGGTGAAGAAAGACCTGTGATCTCGGAGAACTACAG 3537
Db 101 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 120

QY 3538 AACTGCGGTGTTCCCGAAGGCCCGCTGAGACTGGAGTGGCCGCGATAGGATGCC 3597
Db 121 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 140

QY 3598 CCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCATCCCTTGCACGAGAGAGGATC 3657
Db 141 ProProValPheTyrlsLysLysAspAsnGluThrIleProCysThrArgGluArgIle 160

QY 3658 AGTATGCCAGGACACACAGGATGATCGCTGCTTCATTCAGCCAGCAAGAAA- 3714
Db 161 SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLys 179

RESULT 5
ABBI0509
ID ABB10509 standard; Protein; 159 AA.
XX
AC ABB10509;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 817.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX
OS Homo sapiens.
XX
PN W0200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228944P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-231415P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249246P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.

PR	14-SEP-2000;	2000US-0233297;
PR	14-SEP-2000;	2000US-0233298;
PR	14-SEP-2000;	2000US-0233299;
PR	14-SEP-2000;	2000US-0233300;
PR	14-SEP-2000;	2000US-0234001;
PR	14-SEP-2000;	2000US-0234002;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	14-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0234223;
PR	21-SEP-2000;	2000US-0234274;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	26-SEP-2000;	2000US-0234984;
PR	27-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236327;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	02-OCT-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	02-OCT-2000;	2000US-0237040;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;
PR	01-NOV-2000;	2000US-0244617;
PR	08-NOV-2000;	2000US-0246474;
PR	08-NOV-2000;	2000US-0246475;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246525;
PR	08-NOV-2000;	2000US-0246526;
PR	08-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246528;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246609;
PR	08-NOV-2000;	2000US-0246610;
PR	08-NOV-2000;	2000US-0246611;
PR	08-NOV-2000;	2000US-0246613;
PR	17-NOV-2000;	2000US-0249207;
PR	17-NOV-2000;	2000US-0249208;
PR	17-NOV-2000;	2000US-0249214;
PR	17-NOV-2000;	2000US-0249215;
PR	17-NOV-2000;	2000US-0249216;
PR	17-NOV-2000;	2000US-0249217;
PR	17-NOV-2000;	2000US-0249211;
PR	17-NOV-2000;	2000US-0249212;
PR	17-NOV-2000;	2000US-0249213;
PR	17-NOV-2000;	2000US-0249214;
PR	17-NOV-2000;	2000US-0249245;
PR	17-NOV-2000;	2000US-0249264;
PR	17-NOV-2000;	2000US-0249265;
PR	17-NOV-2000;	2000US-0249265;
PR	17-NOV-2000;	2000US-0249297;
PR	17-NOV-2000;	2000US-0249299;
PR	17-NOV-2000;	2000US-0249300;
PR	01-DEC-2000;	2000US-0250160;
PR	01-DEC-2000;	2000US-0250391;

PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-)	HUMAN GENOME SCI INC.	
XX	Rosen CA,	Barash SC, Ruben SM;	
XX	WPI;	2001-457725/49.	
DR	N-PSDB;	AAS28841.	
DR			
PT	Isolated novel immunoglobulin polypeptide for monitoring the presence		
PT	and progression of diseases and for diagnosis		
XX	Claim 11; SEQ ID No 198; 551pp; English.		
XX	Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the		
CC	invention. The polypeptides and their associated polynucleotides can be		
CC	used to diagnose a pathological condition or a susceptibility to a		
CC	pathological condition in a subject by determining the presence or		
CC	absence of a mutation in a DNA sequence or determining the presence or		
CC	amount of expression of the protein. Alternatively the identification of		
CC	a binding partner to a sequence allows determination of changes in		
CC	protein activity. The sequences can be used as research tools for		
CC	receptors or other signal transduction pathway proteins that interact		
CC	with the polypeptides of the invention and can be used to treat, prevent,		
CC	or diagnose various types of disorders such as neurological disorders,		
CC	cardiovascular disorders, gastrointestinal disorders, reproductive		
CC	disorders, immune system disorders, renal disorders, muscular disorders,		
CC	pulmonary disorders, proliferative disorders and cancer.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SQ	Sequence	159 AA;	
Alignment Scores:			
Pred. No.:	4,31e-95	Length:	159
Score:	113.00	Matches:	113
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.59%	Indels:	0
DB:	22	Gaps:	0
US-09-818-990B-1 (1-3963) x AAU18053 (1-159)			
Qy	3424	GGGACCTATAAGTCATCGCTACCAACAAACCGGCAGAAATTCCTTTAGTCTGGAGCTC	3483
Db	8	GlyThryrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeu	27
Qy	3484	TCTGTAGTACCAAGAGGTTGAAGAAGCACCCTGTGTACTCTGGAGAACTACAGAACTGC	3543
Db	28	SerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys	47
Qy	3544	GGTGTTCCTCCGAGCCACCCTGCAGACTGGAGTGGCGCTGATAGGCATGCCCCACCT	3603
Db	48	GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProPro	67
Qy	3604	GTGTTCTACTGGAAAGAACAATGAGACCATCCCTTGCACACAGAGAGGATCAGTATG	3663
Db	68	ValPheYrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet	87
Qy	3664	CACGAGACACACAGGGTATGCCTTCTCATTCAGCCAGCCAAATCAGACGCT	3723
Db	88	HisGlnAspThrTrpGlyTrpAlaCysLeuLeuIleGlnProAlaLysLysSAspAla	107

QY 3724 GGATGTCACACGTGTGTCAGCCAGAAATGAAGCGGCATC 3762
|||||
Db 108 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle 120

RESULT 7
AAB40875
ID AAB40875 standard; Protein; 130 AA.
XX AAB40875;
XX
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF639 polypeptide sequence SEQ ID NO:1278.
XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.

OS WO2000058473-A2.

PN 05-OCT-2000.

PD

PF 31-MAR-2000; 2000WO-US08621.

PP 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC75084.

XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 1115-1116; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 130 AA;

Alignment Scores:

Pred. No.: 154e-65 Length: 130
Score: 81.00 Matches: 81
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.16% Indels: 0
DB: 21 Gaps: 0

US-09-818-990B-1 (1-3963) x AAB40875 (1-130)

QY 2542 CTGCTAGAGTGCACCATCCATCCAGGATTCAGGAGAAATAACAAGTCT 2601
|||||

Db 43 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer 62
|||||

QY 2602 CCTCACCCAGTGAATGATGATTAACATTCGTGAACCTAAGAACCGAGTTCGAGACTTG 2661
|||||

Db 63 ProGlnProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeu 82
|||||

QY 2662 GGGAAAAAATAACTTTCAGTGTGTGACACCAACCCAGCAGGATTCACAAATTCAGC 2721
|||||

Db 83 GlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGluThrLysIleSerSer 102
|||||

QY 2722 TTTGACGACAGGCTGATGAATGAATAGAGTTTCGGCTTGAAGCTACTCTGTTGATGAA 2781
|||||

Db 103 PheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGlu 122
|||||

QY 2782 TCA 2784
|||

Db 123 Ser 123

RESULT 8

ABP09582

ID ABP09582 standard; Protein; 50 AA.

XX AC ABP09582;

XX 24-JUN-2002 (first entry)

XX Human ORFX protein sequence SEQ ID NO:19146.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis.

XX OS Homo sapiens.

XX PN WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.

XX 30-MAY-2000; 2000US-206132P.

XX 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABN25334.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders
XX
XX Disclosure; SEQ ID 19146; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ASN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 50 AA;

Alignment Scores:
Pred. No.: 3,66e-21 Length: 50
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.51% Indels: 0
DB: 23 Gaps: 0

US-09-818-990b-1 (1-3963) x ABP09582 (1-50)

QY 2122 AAGCAGGTGAAGCTCCTCATCACAGACGTTGCGTGGCGGAGATTTCTTC 2181
Db 17 LysGlnValLysAlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePhe 36
QY 2182 CCTCCACGACACCCGACGACACTGTGCCCCCTTCC 2220
Db 37 ProSerThrAsnThrThrAlaAlaThrValAlaProSer 49

RESULT 9
AAB42762
ID AAB42762 standard; Protein: 433 AA.

XX AAB42762;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2526 polypeptide sequence SEQ ID NO:5052.

XX Human; open reading frame; ORFX: detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC76971.

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PT
XX
XX Claim 11; Page 4239-4240; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 433 AA;

Alignment Scores:
Pred. No.: 2,19e-08 Length: 433
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
-Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 21 Gaps: 0

US-09-818-990b-1 (1-3963) x AAB42762 (1-433)

QY 3034 GATGACGATGGCAACATACACCATCATGCGACCAACCCCGGGGAGAAATCAGCTGT 3090
Db 118 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 136

RESULT 10
AAB53464
ID AAB53464 standard; Protein: 544 AA.
XX

AC AAB53464;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:1004.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
OS Homo sapiens.
XX
PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR N-PSDB; AAC98221.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11; Page 1585-1587; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
XX invention.
SQ Sequence 544 AA;
Alignment Scores:
Pred. No.: 2.12e-08 Length: 544
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 21 Gaps: 0
US-09-818-990B-1 (1-3963) x AAB53464 (1-544)
QY 3034 GATGACGATGCACTACACATCATGGCAGCCACCCAGGGGAGAGATCAGCTGT 3090
|||||
Db 229 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 247
|||||
RESULT 11
AAU01767
ID AAU01767 standard; Protein; 43 AA.

XX AAU01767;
AC 18-JUL-2001 (first entry)
XX
DT Human secreted protein #46.
XX
DE Human; secreted protein; immunogen; antibody; diagnosis;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW fungal infection; corneal infection; wound healing; cell culture;
KW epithelial cell proliferation; skin ageing; transplantation;
KW tissue regeneration; chemotaxis; food additive.
XX
OS Homo sapiens.
XX
PN WO200123546-A1.
XX
PD 05-APR-2001.
XX
PF 26-SEP-2000; 2000WO-US26323.
XX
PR 27-SEP-1999; 99US-0155805.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben S, Komatsoulis GA;
XX
DR WPI; 2001-266150/27.
DR N-PSDB; AAS02397.
XX
PT Nucleic acids encoding 37 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Disclosure; Page 457; 494pp; English.
XX
CC The sequence represents a human secreted protein of the invention. The
CC polynucleotides, polypeptides and antibodies raised against them are used
CC to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polynucleotides and antibodies are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. The antibodies
CC can also be used in alleviating symptoms associated with the disorders
CC and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
XX
SQ Sequence 43 AA;
Alignment Scores:
Pred. No.: 1.85e-05 Length: 43
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 22 Gaps: 0
US-09-818-990B-1 (1-3963) x AAU01767 (1-43)
QY 3739 TCAGCCAAGATGAAGCGGCATCGTGTGTCGACTGCCAGCTGGAT 3786
|||||

```
Db      22 SerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 37
RESULT 12
AAU01765
ID      AAU01765 standard; Protein; 44 AA.
XX
AC      AAU01765;
XX
DT      18-JUL-2001 (first entry)
XX
DE      Human secreted protein #44.
XX
KW      Human; secreted protein; immunogen; antibody; diagnosis;
KW      rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW      cardiovascular disorder; cerebrovascular disorder; cerebral
KW      angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW      fungal infection; corneal infection; wound healing; cell culture;
KW      epithelial cell proliferation; skin ageing; transplantation;
KW      tissue regeneration; chemotaxis; food additive.
XX
OS      Homo sapiens.
XX
PN      WO200123546-A1.
XX
PD      05-APR-2001.
XX
PF      26-SEP-2000; 2000WO-US26323.
XX
PR      27-SEP-1999; 99US-0155805.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Rosen CA, Ruben S, Komatsoulis GA;
XX
DR      WPI; 2001-266150/27.
XX
DR      N-PSDB; AAS02397.
XX
XX
XX      Nucleic acids encoding 37 human secreted polypeptides, useful for
XX      preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX      disease and diabetic retinopathy -
XX
PS      Disclosure; Page 456-457; 494pp; English.
XX
XX      The sequence represents a human secreted protein of the invention. The
XX      polynucleotides, polypeptides and antibodies raised against them are used
XX      to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX      rabbits, goats, horses, cats, dogs, chickens or sheep. The
XX      polynucleotides and antibodies are also used in diagnosing a pathological
XX      condition or susceptibility to a pathological condition. The antibodies
XX      can also be used in alleviating symptoms associated with the disorders
XX      and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX      immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX      include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX      disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX      e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX      angiogenesis, nervous system disorders e.g. Alzheimer's disease,
XX      infections caused by bacteria, viruses and fungi and ocular disorders
XX      e.g. corneal infection. The polypeptides can also be used to aid wound
XX      healing and epithelial cell proliferation, to prevent skin aging due to
XX      sunburn, to maintain organs before transplantation, for supporting cell
XX      culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX      polypeptides can also be used as a food additive or preservative to
XX      increase or decrease storage capabilities.
XX
SQ      Sequence 44 AA;
Alignment Scores:
Pred. No.: 1.84e-05 Length: 44
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 22 Gaps: 0
```

```
US-09-818-990B-1 (1-3963) x AAU01765 (1-44)
QY      3739 TCACCCAGATGAAGCGGCATCGTGTGCGACTGCCAGGCTGGAT 3786
      |||||||
Db      28 SerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 43
RESULT 13
AAU01763
ID      AAU01763 standard; Protein; 49 AA.
XX
AC      AAU01763;
XX
DT      18-JUL-2001 (first entry)
XX
DE      Human secreted protein #42.
XX
KW      Human; secreted protein; immunogen; antibody; diagnosis;
KW      rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW      cardiovascular disorder; cerebrovascular disorder; cerebral
KW      angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW      fungal infection; corneal infection; wound healing; cell culture;
KW      epithelial cell proliferation; skin ageing; transplantation;
KW      tissue regeneration; chemotaxis; food additive.
XX
OS      Homo sapiens.
XX
PN      WO200123546-A1.
XX
PD      05-APR-2001.
XX
PF      26-SEP-2000; 2000WO-US26323.
XX
PR      27-SEP-1999; 99US-0155805.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Rosen CA, Ruben S, Komatsoulis GA;
XX
DR      WPI; 2001-266150/27.
XX
DR      N-PSDB; AAS02397.
XX
XX
XX      Nucleic acids encoding 37 human secreted polypeptides, useful for
XX      preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX      disease and diabetic retinopathy -
XX
PS      Disclosure; Page 456; 494pp; English.
XX
XX      The sequence represents a human secreted protein of the invention. The
XX      polynucleotides, polypeptides and antibodies raised against them are used
XX      to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX      rabbits, goats, horses, cats, dogs, chickens or sheep. The
XX      polynucleotides and antibodies are also used in diagnosing a pathological
XX      condition or susceptibility to a pathological condition. The antibodies
XX      can also be used in alleviating symptoms associated with the disorders
XX      and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX      immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX      include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX      disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX      e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX      angiogenesis, nervous system disorders e.g. Alzheimer's disease,
XX      infections caused by bacteria, viruses and fungi and ocular disorders
XX      e.g. corneal infection. The polypeptides can also be used to aid wound
XX      healing and epithelial cell proliferation, to prevent skin aging due to
XX      sunburn, to maintain organs before transplantation, for supporting cell
XX      culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX      polypeptides can also be used as a food additive or preservative to
XX      increase or decrease storage capabilities.
XX
SQ      Sequence 49 AA;
Alignment Scores:
Pred. No.: 1.81e-05 Length: 49
```

Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1-22% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x AAU01763 (1-49)

QY 3739 TCAGCCAGGAATGAAGCGGCATCGTGTGTCGACATGCCAGGTGGAT 3786

|||||
Db 28 SerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 43

RESULT 14

AAU06795

ID AAU06795 standard; Protein: 104 AA.

XX

AC AAU06795;

XX

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 1003.

XX

KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW neutropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation.

XX

OS Homo sapiens.

XX

PN WO20015339-A2.

XX

PD 02-AUG-2001.

XX

XX 25-JAN-2001; 2001WO-US02723.

PF

XX 25-JAN-2000; 2000US-0491404.

XX 15-SEP-2000; 2000US-0563870.

PR 06-NOV-2000; 2000US-0707351.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;

PI Liu C, Asundi V, Zhou P, Werhman T;

XX

DR WPI: 2001-465571/50.

DR N-PSDB; AAH94470.

XX

PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -

XX

PS Example 4; Page 577-578; 715pp; English.

XX

CC The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a polypeptide encoded by a cDNA assembled using
CC an expressed sequence tag (EST) found to be expressed in human
CC foetal tissue cDNA libraries.

XX Sequence 104 AA;

Alignment Scores:

Pred. No.: 0.00113 Length: 104
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.06% Indels: 0

DB: 22 Gaps: 0
US-09-818-990B-1 (1-3963) x AAU06795 (1-104)

QY 3034 GATGACGATGGCACTACACCATCATGTGCAGGCAACCCCCAG 3075
|||||
Db 73 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGln 86

RESULT 15

ABP07380

ID ABP07380 standard; Protein: 162 AA.

XX

AC ABP07380;

XX

DT 25-JUN-2002 (first entry)

DE

XX Human ORFX protein sequence SEQ ID NO:14742.

XX

KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.

XX

OS Homo sapiens.

XX

PN WO200192523-A2.

XX

PD 06-DEC-2001.

XX

XX 29-MAY-2001; 2001WO-US10836.

XX

XX 30-MAY-2000; 2000US-206132P.

PR 29-AUG-2000; 2000US-228716P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach MD;

XX

DR WPI; 2002-106308/14.

DR

DR N-PSDB; ABN23132.

XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -

XX

PS Disclosure; SEQ ID 14742; 1037pp; English.

XX

CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is I-I1491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.

CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 162 AA;

Alignment Scores:
Pred. No.: 0.00106 Length: 162
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.06% Indels: 0
DB: 23 Gaps: 0

US-09-818-990B-1 (1-3963) x ABP07380 (1-162)

Oy 1429 GATTCTCCAGATTTTAGGATTTTACAGAAAAACCTCGATCC 1470
Db 127 AspSerProAspPheArgIleLeuGlnLysLysProArgSer 140

Search completed: November 30, 2002, 18:43:24
Job time : 103 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:41:29 ; Search time 23 Seconds
(without alignments)
10139.395 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 1316
Sequence: 1 atycaagacgacagcataga.....tggagagtgtgaacttta 3963

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/Cgn2_1/USPTO.spool/US09818990/runat_26112002_093223_22098/app_query_fasta_1.4103
-DB-Issued_Patents_AA_QFMT-fastan_SUFFIX=oligo.rai -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09818990.ecgn_1.1.20 -runat_26112002_093223_22098 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /Cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /Cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /Cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /Cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /Cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /Cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
C 1	8	0.6 192 4	US-09-149-476-477 Sequence 477, Appl
C 2	8	0.6 305 4	US-09-500-569-2 Sequence 2, Appli
C 3	8	0.6 352 2	US-08-440-845D-5 Sequence 5, Appli
C 4	8	0.6 352 4	US-08-868-458-5 Sequence 5, Appli
C 5	8	0.6 352 4	US-09-134-262-2 Sequence 2, Appli
C 6	8	0.6 352 4	US-09-359-070-2 Sequence 2, Appli
C 7	8	0.6 362 1	US-08-464-523B-32 Sequence 32, Appl
C 8	8	0.6 362 2	US-08-948-176-25 Sequence 25, Appl
C 9	8	0.6 362 2	US-08-440-845D-6 Sequence 6, Appli
C 10	8	0.6 362 4	US-08-868-458-6 Sequence 6, Appli
C 11	8	0.6 362 4	US-09-303-592-2 Sequence 2, Appli
C 12	8	0.6 362 4	US-09-303-592-4 Sequence 4, Appli

c 86	7	0.5	133	3	US-08-513-974B-329	Sequence 329, App	c 159	7	0.5	310	4	US-09-528-784A-19	Sequence 19, Appl
c 87	7	0.5	137	2	US-08-757-036-4	Sequence 4, Appli	c 160	7	0.5	312	1	US-08-094-128A-27	Sequence 27, Appl
c 88	7	0.5	137	4	US-08-936-165A-431	Sequence 431, App	c 161	7	0.5	312	1	US-08-455-674-27	Sequence 27, Appl
c 89	7	0.5	139	3	US-08-950-720A-15	Sequence 15, Appl	c 162	7	0.5	312	1	US-08-455-992-27	Sequence 27, Appl
c 90	7	0.5	139	3	US-08-991-690-2	Sequence 2, Appli	c 163	7	0.5	312	1	US-08-455-972-27	Sequence 27, Appl
c 91	7	0.5	139	4	US-09-599-364A-2	Sequence 2, Appli	c 164	7	0.5	312	5	PCT-US92-00652-27	Sequence 27, Appl
c 92	7	0.5	144	3	US-09-134-001C-4761	Sequence 4761, Ap	c 165	7	0.5	314	4	US-08-486-036A-4	Sequence 4, Appli
c 93	7	0.5	144	3	US-08-513-974B-334	Sequence 334, App	c 166	7	0.5	314	4	US-09-005-298-4	Sequence 4, Appli
c 94	7	0.5	147	4	US-08-540-650B-13	Sequence 13, Appl	c 167	7	0.5	314	4	US-08-768-619-4	Sequence 4, Appli
c 95	7	0.5	148	2	US-08-858-767-31	Sequence 31, Appl	c 168	7	0.5	314	5	PCT-US96-09848-4	Sequence 4, Appli
c 96	7	0.5	148	2	US-08-863-028-31	Sequence 31, Appl	c 169	7	0.5	316	1	US-08-482-282B-4	Sequence 4, Appli
c 97	7	0.5	149	3	US-08-834-224-4	Sequence 4, Appli	c 170	7	0.5	319	2	US-08-149-975A-2	Sequence 2, Appli
c 98	7	0.5	156	4	US-09-134-001C-5423	Sequence 5423, Ap	c 171	7	0.5	321	2	US-08-592-126-143	Sequence 143, App
c 99	7	0.5	174	4	US-09-174-465D-2	Sequence 2, Appli	c 172	7	0.5	337	1	US-08-153-848-46	Sequence 46, Appl
c 100	7	0.5	174	4	US-09-072-596-281	Sequence 281, App	c 173	7	0.5	337	3	US-09-299-843A-46	Sequence 46, Appl
c 101	7	0.5	178	3	US-09-053-197A-24	Sequence 24, Appl	c 174	7	0.5	337	4	US-09-088-337B-46	Sequence 46, Appl
c 102	7	0.5	178	4	US-09-085-761A-24	Sequence 24, Appl	c 175	7	0.5	337	5	PCT-US93-11153-46	Sequence 46, Appl
c 103	7	0.5	185	4	US-09-134-001C-5480	Sequence 5480, Ap	c 176	7	0.5	337	5	US-08-445-515-58	Sequence 58, Appl
c 104	7	0.5	197	3	US-08-737-248-16	Sequence 16, Appl	c 177	7	0.5	348	3	US-08-445-515-56	Sequence 56, Appl
c 105	7	0.5	203	4	US-09-124-141-9	Sequence 9, Appli	c 178	7	0.5	351	1	US-08-197-792-39	Sequence 39, Appl
c 106	7	0.5	205	4	US-09-385-259-3	Sequence 3, Appli	c 179	7	0.5	351	1	US-08-459-850-39	Sequence 39, Appl
c 107	7	0.5	205	4	US-09-191-468-55	Sequence 55, Appl	c 180	7	0.5	351	1	US-08-459-214-39	Sequence 39, Appl
c 108	7	0.5	205	4	US-09-191-468-57	Sequence 57, Appl	c 181	7	0.5	352	1	US-08-202-056-3	Sequence 3, Appli
c 109	7	0.5	205	4	US-09-191-468-59	Sequence 59, Appl	c 182	7	0.5	352	1	US-08-076-093A-4	Sequence 4, Appli
c 110	7	0.5	205	4	US-09-191-468-61	Sequence 61, Appl	c 183	7	0.5	352	1	US-08-450-393A-6	Sequence 6, Appli
c 111	7	0.5	205	4	US-09-191-468-63	Sequence 63, Appl	c 184	7	0.5	352	1	US-08-701-265-4	Sequence 4, Appli
c 112	7	0.5	205	4	US-09-191-468-68	Sequence 68, Appl	c 185	7	0.5	352	2	US-08-284-586-4	Sequence 4, Appli
c 113	7	0.5	205	4	US-09-191-468-70	Sequence 70, Appl	c 186	7	0.5	352	2	US-08-805-478-4	Sequence 4, Appli
c 114	7	0.5	205	4	US-09-645-370-3	Sequence 3, Appli	c 187	7	0.5	352	2	US-08-802-627A-4	Sequence 4, Appli
c 115	7	0.5	207	3	US-08-996-338-23	Sequence 23, Appl	c 188	7	0.5	352	2	US-08-801-238-4	Sequence 4, Appli
c 116	7	0.5	213	4	US-09-005-298-22	Sequence 22, Appl	c 189	7	0.5	352	2	US-08-801-228-4	Sequence 4, Appli
c 117	7	0.5	213	4	US-08-768-619-22	Sequence 22, Appl	c 190	7	0.5	352	3	US-09-104-296-4	Sequence 4, Appli
c 118	7	0.5	213	5	PCT-US96-09848-21	Sequence 21, Appl	c 191	7	0.5	352	4	US-08-446-669-6	Sequence 6, Appli
c 119	7	0.5	215	4	US-09-005-298-38	Sequence 38, Appl	c 192	7	0.5	352	4	US-08-446-669-6	Sequence 6, Appli
c 120	7	0.5	215	4	US-08-768-619-38	Sequence 38, Appl	c 193	7	0.5	352	5	PCT-US95-00476-6	Sequence 6, Appli
c 121	7	0.5	220	4	US-09-134-001C-5073	Sequence 5073, Ap	c 194	7	0.5	354	2	US-08-700-013B-9	Sequence 9, Appli
c 122	7	0.5	227	4	US-09-308-003-48	Sequence 48, Appl	c 195	7	0.5	355	2	US-08-440-845D-25	Sequence 25, Appl
c 123	7	0.5	232	4	US-09-149-476-623	Sequence 623, App	c 196	7	0.5	356	4	US-09-005-298-25	Sequence 25, Appl
c 124	7	0.5	232	4	US-09-149-476-623	Sequence 623, App	c 197	7	0.5	356	4	US-08-768-619-25	Sequence 25, Appl
c 125	7	0.5	241	4	US-09-005-298-46	Sequence 46, Appl	c 198	7	0.5	356	5	PCT-US96-09848-24	Sequence 24, Appl
c 126	7	0.5	241	4	US-08-768-619-46	Sequence 46, Appl	c 199	7	0.5	358	3	US-09-041-545-2	Sequence 2, Appli
c 127	7	0.5	242	4	US-09-134-001C-5154	Sequence 5154, Ap	c 200	7	0.5	358	3	US-09-327-925-2	Sequence 2, Appli
c 128	7	0.5	244	4	US-08-918-148-77	Sequence 77, Appl	c 201	7	0.5	362	4	US-08-637-670-35	Sequence 35, Appl
c 129	7	0.5	245	4	US-08-918-148-76	Sequence 76, Appl	c 202	7	0.5	364	1	US-08-197-792-29	Sequence 29, Appl
c 130	7	0.5	249	1	US-08-450-257-57	Sequence 57, Appl	c 203	7	0.5	364	1	US-08-459-850-29	Sequence 29, Appl
c 131	7	0.5	249	1	US-08-450-246-57	Sequence 57, Appl	c 204	7	0.5	364	1	US-08-459-214-29	Sequence 29, Appl
c 132	7	0.5	249	1	US-08-450-098-57	Sequence 57, Appl	c 205	7	0.5	366	4	US-09-134-001C-3198	Sequence 3198, Ap
c 133	7	0.5	249	1	US-08-451-233-57	Sequence 57, Appl	c 206	7	0.5	367	1	US-08-075-533-2	Sequence 2, Appli
c 134	7	0.5	249	1	US-08-450-236-57	Sequence 57, Appl	c 207	7	0.5	367	1	US-08-075-533-4	Sequence 4, Appli
c 135	7	0.5	249	4	US-08-235-403-57	Sequence 57, Appl	c 208	7	0.5	367	2	US-08-948-176-2	Sequence 2, Appli
c 136	7	0.5	254	4	US-09-199-637A-241	Sequence 241, App	c 209	7	0.5	367	2	US-08-948-176-4	Sequence 4, Appli
c 137	7	0.5	259	4	US-09-253-316-27	Sequence 27, Appl	c 210	7	0.5	367	4	US-08-845-258-20	Sequence 20, Appl
c 138	7	0.5	259	6	5212074-3	Patent No. 5212074	c 211	7	0.5	367	4	US-08-845-258-20	Sequence 20, Appl
c 139	7	0.5	263	4	US-08-845-258-18	Sequence 18, Appl	c 212	7	0.5	367	4	US-08-723-142A-20	Sequence 20, Appl
c 140	7	0.5	263	4	US-08-990-571-18	Sequence 18, Appl	c 213	7	0.5	367	5	US-09-528-784A-20	Sequence 20, Appl
c 141	7	0.5	263	4	US-08-723-142A-18	Sequence 18, Appl	c 214	7	0.5	367	5	PCT-US91-09160-2	Sequence 2, Appli
c 142	7	0.5	263	4	US-09-528-784A-18	Sequence 18, Appl	c 215	7	0.5	367	5	PCT-US91-09160-4	Sequence 4, Appli
c 143	7	0.5	277	2	US-08-147-784-2	Sequence 2, Appli	c 216	7	0.5	371	2	US-08-928-692-20	Sequence 20, Appl
c 144	7	0.5	277	4	US-08-195-967-2	Sequence 2, Appli	c 217	7	0.5	371	4	US-09-339-972-20	Sequence 20, Appl
c 145	7	0.5	277	4	US-09-006-353A-12	Sequence 12, Appl	c 218	7	0.5	375	1	US-08-464-523B-22	Sequence 22, Appl
c 146	7	0.5	277	4	US-08-472-940-2	Sequence 2, Appli	c 219	7	0.5	376	2	US-08-883-534-5	Sequence 5, Appli
c 147	7	0.5	277	4	US-09-573-986-12	Sequence 12, Appl	c 220	7	0.5	376	3	US-09-204-767-5	Sequence 5, Appli
c 148	7	0.5	277	6	5164481-2	Patent No. 5164481	c 221	7	0.5	377	4	US-09-269-137-4	Sequence 4, Appli
c 149	7	0.5	279	4	US-09-181-183-34	Sequence 34, Appl	c 222	7	0.5	384	3	US-09-071-434-3	Sequence 3, Appli
c 150	7	0.5	279	4	US-09-280-040-34	Sequence 34, Appl	c 223	7	0.5	384	4	US-09-334-601-8	Sequence 8, Appli
c 151	7	0.5	279	4	US-09-277-700-34	Sequence 34, Appl	c 224	7	0.5	390	4	US-09-308-003-12	Sequence 12, Appl
c 152	7	0.5	280	4	US-09-194-146-10	Sequence 10, Appl	c 225	7	0.5	394	3	US-08-673-814-6	Sequence 6, Appli
c 153	7	0.5	282	4	US-09-404-879A-393	Sequence 393, App	c 226	7	0.5	394	4	US-09-115-824-6	Sequence 6, Appli
c 154	7	0.5	309	3	US-08-996-338-22	Sequence 32, Appl	c 227	7	0.5	396	4	US-09-134-001C-4580	Sequence 4580, Ap
c 155	7	0.5	309	4	US-09-404-879A-392	Sequence 392, App	c 228	7	0.5	396	1	US-08-486-036A-2	Sequence 2, Appli
c 156	7	0.5	310	4	US-08-845-258-19	Sequence 19, Appl	c 229	7	0.5	398	4	US-09-005-298-2	Sequence 2, Appli
c 157	7	0.5	310	4	US-08-990-571-19	Sequence 19, Appl	c 230	7	0.5	398	4	US-08-768-619-2	Sequence 2, Appli
c 158	7	0.5	310	4	US-08-723-142A-19	Sequence 19, Appl	c 231	7	0.5	398	5	PCT-US96-09848-2	Sequence 2, Appli

c 232	7	0.5	400	1	US-08-482-282B-2	Sequence 2, Appli	c 305	7	0.5	632	4	US-09-232-197-34	Sequence 34, Appl
c 233	7	0.5	400	4	US-09-005-298-33	Sequence 33, Appl	c 306	7	0.5	632	4	US-09-232-197-35	Sequence 35, Appl
c 234	7	0.5	400	4	US-08-768-619-33	Sequence 33, Appl	c 307	7	0.5	632	4	US-09-232-197-39	Sequence 39, Appl
c 235	7	0.5	401	4	US-09-005-298-44	Sequence 44, Appl	c 308	7	0.5	632	4	US-09-232-201-34	Sequence 34, Appl
c 236	7	0.5	401	4	US-08-768-619-44	Sequence 44, Appl	c 309	7	0.5	632	4	US-09-232-201-35	Sequence 35, Appl
c 237	7	0.5	405	2	US-08-700-013B-6	Sequence 6, Appli	c 310	7	0.5	632	4	US-09-232-201-39	Sequence 39, Appl
c 238	7	0.5	409	3	US-09-075-215A-17	Sequence 17, Appl	c 311	7	0.5	632	5	PCT-US93-01959-10	Sequence 10, Appl
c 239	7	0.5	410	2	US-08-472-666-2	Sequence 2, Appli	c 312	7	0.5	642	1	US-08-706-936-2	Sequence 2, Appli
c 240	7	0.5	414	5	PCT-US96-07615-2	Sequence 2, Appli	c 313	7	0.5	642	1	US-08-706-936-3	Sequence 3, Appli
c 241	7	0.5	410	5	PCT-US92-06840-2	Sequence 2, Appli	c 314	7	0.5	643	2	US-08-616-844-39	Sequence 39, Appl
c 242	7	0.5	420	2	US-08-592-126-142	Sequence 142, App	c 315	7	0.5	643	2	US-08-599-654-39	Sequence 39, Appl
c 243	7	0.5	422	1	US-08-132-649-4	Sequence 4, Appli	c 316	7	0.5	643	3	US-08-944-868A-39	Sequence 39, Appl
c 244	7	0.5	422	3	US-08-767-579-4	Sequence 4, Appli	c 317	7	0.5	643	3	US-08-944-423A-39	Sequence 39, Appl
c 245	7	0.5	425	2	US-08-951-148-3	Sequence 3, Appli	c 318	7	0.5	643	3	US-08-944-496-39	Sequence 39, Appl
c 246	7	0.5	425	2	US-09-165-234-3	Sequence 3, Appli	c 319	7	0.5	643	4	US-09-232-200-27	Sequence 27, Appl
c 247	7	0.5	425	3	US-09-274-570-3	Sequence 3, Appli	c 320	7	0.5	643	4	US-09-232-200-41	Sequence 41, Appl
c 248	7	0.5	427	3	US-08-705-771-16	Sequence 16, Appl	c 321	7	0.5	643	4	US-09-232-200-42	Sequence 42, Appl
c 249	7	0.5	448	1	US-08-570-157-3	Sequence 3, Appli	c 322	7	0.5	643	4	US-09-232-200-45	Sequence 45, Appl
c 250	7	0.5	448	4	US-09-076-510-3	Sequence 3, Appli	c 323	7	0.5	643	4	US-09-232-200-53	Sequence 53, Appl
c 251	7	0.5	455	2	US-08-588-983-7	Sequence 7, Appli	c 324	7	0.5	643	4	US-09-232-200-71	Sequence 71, Appl
c 252	7	0.5	455	2	US-08-588-976-7	Sequence 7, Appli	c 325	7	0.5	643	4	US-09-232-197-27	Sequence 27, Appl
c 253	7	0.5	458	4	US-08-618-485B-1	Sequence 1, Appli	c 326	7	0.5	643	4	US-09-232-197-41	Sequence 41, Appl
c 254	7	0.5	458	6	5177002-1	Patent No. 5177002	c 327	7	0.5	643	4	US-09-232-197-42	Sequence 42, Appl
c 255	7	0.5	458	6	5177002-2	Patent No. 5177002	c 328	7	0.5	643	4	US-09-232-197-45	Sequence 45, Appl
c 256	7	0.5	459	4	US-09-071-709-1	Sequence 1, Appli	c 329	7	0.5	643	4	US-09-232-197-53	Sequence 53, Appl
c 257	7	0.5	460	3	US-09-199-637A-132	Sequence 132, App	c 330	7	0.5	643	4	US-09-232-197-71	Sequence 71, Appl
c 258	7	0.5	465	3	US-08-788-231A-2	Sequence 2, Appli	c 331	7	0.5	643	4	US-09-232-201-27	Sequence 27, Appl
c 259	7	0.5	474	1	US-08-222-619-5	Sequence 5, Appli	c 332	7	0.5	643	4	US-09-232-201-41	Sequence 41, Appl
c 260	7	0.5	474	5	PCT-US95-04075-5	Sequence 5, Appli	c 333	7	0.5	643	4	US-09-232-201-42	Sequence 42, Appl
c 261	7	0.5	476	1	US-08-565-655-6	Sequence 6, Appli	c 334	7	0.5	643	4	US-09-232-201-45	Sequence 45, Appl
c 262	7	0.5	485	1	US-08-055-797-2	Sequence 2, Appli	c 335	7	0.5	643	4	US-09-232-201-53	Sequence 53, Appl
c 263	7	0.5	485	1	US-07-914-284A-7	Sequence 7, Appli	c 336	7	0.5	643	4	US-09-232-201-71	Sequence 71, Appl
c 264	7	0.5	517	1	US-08-132-649-2	Sequence 2, Appli	c 337	7	0.5	645	4	US-08-791-115B-6	Sequence 6, Appli
c 265	7	0.5	517	1	US-08-764-343-4	Sequence 4, Appli	c 338	7	0.5	647	4	US-08-485-355B-50	Sequence 50, Appl
c 266	7	0.5	517	3	US-08-767-579-2	Sequence 2, Appli	c 339	7	0.5	647	4	US-09-194-613-3	Sequence 3, Appli
c 267	7	0.5	521	3	US-08-996-338-20	Sequence 20, Appl	c 340	7	0.5	651	2	US-08-492-027A-1	Sequence 1, Appli
c 268	7	0.5	522	4	US-08-961-083-120	Sequence 120, App	c 341	7	0.5	652	2	US-08-426-125-9	Sequence 9, Appli
c 269	7	0.5	529	4	US-09-504-358-22	Sequence 22, Appl	c 342	7	0.5	652	2	US-08-455-355-9	Sequence 9, Appli
c 270	7	0.5	529	4	US-09-954-314-22	Sequence 22, Appl	c 343	7	0.5	655	2	US-08-492-027A-6	Sequence 6, Appli
c 271	7	0.5	533	1	US-08-488-308A-6	Sequence 6, Appli	c 344	7	0.5	666	4	US-08-961-083-2	Sequence 2, Appli
c 272	7	0.5	533	4	US-09-385-259-2	Sequence 2, Appli	c 345	7	0.5	667	4	US-09-071-709-9	Sequence 9, Appli
c 273	7	0.5	533	4	US-09-645-370-2	Sequence 2, Appli	c 346	7	0.5	675	4	US-08-485-355B-52	Sequence 52, Appl
c 274	7	0.5	536	4	US-09-186-930-185	Sequence 185, App	c 347	7	0.5	676	2	US-08-398-590A-40	Sequence 40, Appl
c 275	7	0.5	540	1	US-07-906-930B-2	Sequence 2, Appli	c 348	7	0.5	676	4	US-08-894-997-40	Sequence 40, Appl
c 276	7	0.5	541	1	US-08-604-333-2	Sequence 2, Appli	c 349	7	0.5	680	1	US-08-211-430-2	Sequence 2, Appli
c 277	7	0.5	541	3	US-09-110-618-2	Sequence 2, Appli	c 350	7	0.5	680	3	US-08-761-136-1	Sequence 1, Appli
c 278	7	0.5	541	4	US-09-173-151A-28	Sequence 28, Appl	c 351	7	0.5	682	3	US-08-481-433-6	Sequence 6, Appli
c 279	7	0.5	541	4	US-09-578-178-2	Sequence 2, Appli	c 352	7	0.5	702	2	US-08-867-941-25	Sequence 25, Appl
c 280	7	0.5	546	2	US-08-492-027A-8	Sequence 8, Appli	c 353	7	0.5	702	3	US-08-613-009A-9	Sequence 9, Appli
c 281	7	0.5	549	2	US-08-791-887-2	Sequence 2, Appli	c 354	7	0.5	702	4	US-09-074-658-25	Sequence 25, Appl
c 282	7	0.5	550	4	US-09-387-574-10	Sequence 10, Appl	c 355	7	0.5	702	4	US-08-778-570B-11	Sequence 11, Appl
c 283	7	0.5	550	4	US-09-668-096-10	Sequence 10, Appl	c 356	7	0.5	702	4	US-09-059-584-11	Sequence 11, Appl
c 284	7	0.5	565	4	US-08-906-158A-12	Sequence 12, Appl	c 357	7	0.5	706	4	US-09-059-584-46	Sequence 46, Appl
c 285	7	0.5	572	2	US-08-745-934-3	Sequence 3, Appli	c 358	7	0.5	707	2	US-08-949-941B-2	Sequence 2, Appli
c 286	7	0.5	572	4	US-09-147-009-6	Sequence 6, Appli	c 359	7	0.5	709	4	US-08-778-570B-15	Sequence 15, Appl
c 287	7	0.5	573	2	US-08-745-934-4	Sequence 4, Appli	c 360	7	0.5	709	4	US-09-059-584-15	Sequence 15, Appl
c 288	7	0.5	573	4	US-09-147-009-7	Sequence 7, Appli	c 361	7	0.5	712	4	US-09-059-584-51	Sequence 51, Appl
c 289	7	0.5	575	1	US-07-683-957B-2	Sequence 2, Appli	c 362	7	0.5	713	4	US-09-059-584-49	Sequence 49, Appl
c 290	7	0.5	576	4	US-09-347-801-19	Sequence 19, Appl	c 363	7	0.5	713	4	US-09-059-584-53	Sequence 53, Appl
c 291	7	0.5	582	4	US-09-091-725-17	Sequence 17, Appl	c 364	7	0.5	727	4	US-09-235-451-23	Sequence 23, Appl
c 292	7	0.5	593	4	US-09-124-141-7	Sequence 7, Appli	c 365	7	0.5	742	4	US-08-791-115B-5	Sequence 5, Appli
c 293	7	0.5	623	4	US-09-347-801-4	Sequence 4, Appli	c 366	7	0.5	745	2	US-09-010-928B-28	Sequence 28, Appl
c 294	7	0.5	625	4	US-09-347-801-18	Sequence 18, Appl	c 367	7	0.5	764	4	US-09-235-451-36	Sequence 36, Appl
c 295	7	0.5	627	1	US-08-295-814E-4	Sequence 4, Appli	c 368	7	0.5	787	4	US-09-188-930-334	Sequence 334, App
c 296	7	0.5	627	1	US-08-293-361-4	Sequence 10, Appl	c 369	7	0.5	794	1	US-08-366-276-2	Sequence 2, Appli
c 297	7	0.5	627	4	US-09-343-361-4	Sequence 4, Appli	c 370	7	0.5	794	1	US-08-393-333-2	Sequence 2, Appli
c 298	7	0.5	627	5	PCT-US93-01959-4	Sequence 4, Appli	c 371	7	0.5	794	4	US-09-087-465-10	Sequence 10, Appl
c 299	7	0.5	627	5	PCT-US95-10579-10	Sequence 10, Appl	c 372	7	0.5	797	2	US-08-700-013B-19	Sequence 19, Appl
c 300	7	0.5	632	1	US-08-295-814E-10	Sequence 10, Appl	c 373	7	0.5	797	2	US-08-700-013B-21	Sequence 21, Appl
c 301	7	0.5	632	4	US-09-343-361-10	Sequence 10, Appl	c 374	7	0.5	797	4	US-09-182-728A-2	Sequence 2, Appli
c 302	7	0.5	632	4	US-09-232-200-34	Sequence 34, Appl	c 375	7	0.5	797	4	US-09-191-468-120	Sequence 120, App
c 303	7	0.5	632	4	US-09-232-200-35	Sequence 35, Appl	c 376	7	0.5	797	4	US-09-191-468-122	Sequence 122, App
c 304	7	0.5	632	4	US-09-232-200-39	Sequence 39, Appl	c 377	7	0.5	797	4	US-09-191-468-124	Sequence 124, App

C 378	7	0.5	797	4	US-09-795-232-2	Sequence 2, Appl	451	7	0.5	1226	1	US-08-555-678-2	Sequence 2, Appl
C 379	7	0.5	807	4	US-09-081-345-2	Sequence 2, Appl	452	7	0.5	1226	5	PCT-US95-02275-2	Sequence 2, Appl
C 380	7	0.5	817	1	US-08-381-931B-2	Sequence 2, Appl	453	7	0.5	1231	2	US-08-231-193A-48	Sequence 48, Appl
C 381	7	0.5	823	4	US-09-134-001C-4081	Sequence 4081, Ap	454	7	0.5	1231	3	US-08-486-273A-48	Sequence 48, Appl
C 382	7	0.5	835	2	US-08-968-751-4	Sequence 4, Appl	455	7	0.5	1231	3	US-08-480-474-48	Sequence 48, Appl
C 383	7	0.5	866	3	US-09-040-943-2	Sequence 2, Appl	456	7	0.5	1231	3	US-08-940-086A-48	Sequence 48, Appl
C 384	7	0.5	866	4	US-09-621-855-2	Sequence 2, Appl	457	7	0.5	1231	4	US-08-940-035A-48	Sequence 48, Appl
C 385	7	0.5	870	2	US-09-010-928B-2	Sequence 2, Appl	458	7	0.5	1231	4	US-08-935-105A-48	Sequence 48, Appl
C 386	7	0.5	889	4	US-09-132-316-2	Sequence 2, Appl	459	7	0.5	1231	4	US-09-648-797-48	Sequence 48, Appl
C 387	7	0.5	889	4	US-09-132-316-2	Sequence 2, Appl	460	7	0.5	1236	2	US-08-231-193A-6	Sequence 6, Appl
C 388	7	0.5	907	2	US-09-010-928B-4	Sequence 4, Appl	461	7	0.5	1236	2	US-08-486-273A-6	Sequence 6, Appl
C 389	7	0.5	917	1	US-08-245-295-2	Sequence 2, Appl	462	7	0.5	1236	3	US-08-480-474-6	Sequence 6, Appl
C 390	7	0.5	917	1	US-08-481-130-2	Sequence 2, Appl	463	7	0.5	1236	3	US-08-940-086A-6	Sequence 6, Appl
C 391	7	0.5	917	1	US-08-656-984A-2	Sequence 2, Appl	464	7	0.5	1236	4	US-08-940-035A-6	Sequence 6, Appl
C 392	7	0.5	917	1	US-08-485-604-2	Sequence 2, Appl	465	7	0.5	1236	4	US-08-935-105A-6	Sequence 6, Appl
C 393	7	0.5	917	2	US-08-487-595-2	Sequence 2, Appl	466	7	0.5	1236	4	US-09-648-797-6	Sequence 6, Appl
C 394	7	0.5	918	2	US-08-588-983-14	Sequence 14, Appl	467	7	0.5	1239	1	US-08-026-138E-3	Sequence 3, Appl
C 395	7	0.5	918	2	US-08-588-976-14	Sequence 14, Appl	468	7	0.5	1239	2	US-08-231-193A-52	Sequence 52, Appl
C 396	7	0.5	919	2	US-08-588-983-9	Sequence 9, Appl	469	7	0.5	1239	2	US-08-486-273A-52	Sequence 52, Appl
C 397	7	0.5	919	2	US-08-588-983-12	Sequence 12, Appl	470	7	0.5	1239	3	US-08-480-474-52	Sequence 52, Appl
C 398	7	0.5	919	2	US-08-588-976-9	Sequence 9, Appl	471	7	0.5	1239	3	US-08-940-086A-52	Sequence 52, Appl
C 399	7	0.5	919	2	US-08-588-976-12	Sequence 12, Appl	472	7	0.5	1239	3	US-08-940-035A-52	Sequence 52, Appl
C 400	7	0.5	924	1	US-08-481-130-28	Sequence 28, Appl	473	7	0.5	1239	4	US-08-335-105A-52	Sequence 52, Appl
C 401	7	0.5	924	1	US-08-656-984A-28	Sequence 28, Appl	474	7	0.5	1239	4	US-09-648-797-52	Sequence 52, Appl
C 402	7	0.5	924	1	US-08-485-604-28	Sequence 28, Appl	475	7	0.5	1244	2	US-08-231-193A-46	Sequence 46, Appl
C 403	7	0.5	924	2	US-08-487-595-28	Sequence 28, Appl	476	7	0.5	1244	2	US-08-486-273A-46	Sequence 46, Appl
404	7	0.5	964	4	US-08-484-791-2	Sequence 2, Appl	477	7	0.5	1244	3	US-08-480-474-46	Sequence 46, Appl
405	7	0.5	995	5	PCT-US95-04910-14	Sequence 14, Appl	478	7	0.5	1244	3	US-08-940-086A-46	Sequence 46, Appl
406	7	0.5	1040	4	US-08-961-083-118	Sequence 118, App	479	7	0.5	1244	4	US-08-940-035A-46	Sequence 46, Appl
407	7	0.5	1050	4	US-09-428-711A-16	Sequence 16, Appl	480	7	0.5	1244	4	US-08-935-105A-46	Sequence 46, Appl
408	7	0.5	1051	4	US-09-428-711A-16	Sequence 16, Appl	481	7	0.5	1244	4	US-09-648-797-46	Sequence 46, Appl
409	7	0.5	1063	1	US-08-093-453B-3	Sequence 3, Appl	482	7	0.5	1260	4	US-08-506-296B-21	Sequence 21, Appl
410	7	0.5	1063	1	US-08-127-499A-8	Sequence 8, Appl	483	7	0.5	1279	2	US-08-784-649A-2	Sequence 2, Appl
411	7	0.5	1063	1	US-08-482-847-8	Sequence 8, Appl	484	7	0.5	1280	2	US-08-583-276-19	Sequence 19, Appl
412	7	0.5	1065	4	US-09-412-545-2	Sequence 2, Appl	485	7	0.5	1280	2	US-08-752-447-2	Sequence 2, Appl
413	7	0.5	1134	3	US-08-736-214-2	Sequence 2, Appl	486	7	0.5	1280	4	US-09-316-167-2	Sequence 2, Appl
C 414	7	0.5	1138	4	US-09-580-101A-8	Sequence 8, Appl	487	7	0.5	1280	6	520635A-4	Patent No. 5206352
C 415	7	0.5	1151	4	US-09-590-101A-10	Sequence 10, Appl	488	7	0.5	1360	1	US-09-393-569-2	Sequence 2, Appl
416	7	0.5	1156	3	US-08-996-083-1	Sequence 1, Appl	489	7	0.5	1456	1	US-08-026-138E-8	Sequence 8, Appl
417	7	0.5	1156	4	US-09-429-516-1	Sequence 1, Appl	490	7	0.5	1463	1	US-08-157-005-3	Sequence 3, Appl
418	7	0.5	1156	4	US-09-429-516-3	Sequence 3, Appl	491	7	0.5	1463	1	US-08-747-863-3	Sequence 3, Appl
C 419	7	0.5	1157	1	US-07-876-280-30	Sequence 30, Appl	492	7	0.5	1463	4	US-09-565-86A-3	Sequence 3, Appl
C 420	7	0.5	1157	1	US-07-812-180A-2	Sequence 2, Appl	493	7	0.5	1482	1	US-08-026-138E-2	Sequence 2, Appl
C 421	7	0.5	1157	1	US-08-315-468-2	Sequence 2, Appl	494	7	0.5	1484	4	US-08-264-578-2	Sequence 2, Appl
C 422	7	0.5	1157	4	US-07-941-650A-2	Sequence 2, Appl	495	7	0.5	1523	1	US-09-182-024A-2	Sequence 2, Appl
C 423	7	0.5	1163	1	US-08-173-497-4	Sequence 4, Appl	496	7	0.5	1548	1	US-08-463-092B-7	Sequence 7, Appl
C 424	7	0.5	1163	1	US-08-286-889-4	Sequence 4, Appl	497	7	0.5	1548	2	US-08-460-907B-7	Sequence 7, Appl
C 425	7	0.5	1163	1	US-08-485-618-4	Sequence 4, Appl	498	7	0.5	1651	4	US-09-540-245A-18	Sequence 18, Appl
C 426	7	0.5	1163	1	US-08-362-652-4	Sequence 4, Appl	499	7	0.5	1704	4	US-08-485-355B-40	Sequence 40, Appl
C 427	7	0.5	1163	2	US-08-605-672-4	Sequence 4, Appl	500	7	0.5	1832	3	US-09-335-409-4	Sequence 4, Appl
C 428	7	0.5	1163	2	US-08-482-293A-4	Sequence 4, Appl	501	7	0.5	1832	4	US-09-568-102-4	Sequence 4, Appl
C 429	7	0.5	1163	2	US-08-943-363-4	Sequence 4, Appl	502	7	0.5	1832	4	US-09-567-969-4	Sequence 4, Appl
C 430	7	0.5	1163	2	US-08-476-062A-44	Sequence 44, Appl	503	7	0.5	1832	4	US-09-568-480-4	Sequence 4, Appl
C 431	7	0.5	1163	4	US-09-153-043-4	Sequence 4, Appl	504	7	0.5	1832	4	US-09-568-486-4	Sequence 4, Appl
C 432	7	0.5	1163	4	US-09-688-307A-4	Sequence 4, Appl	505	7	0.5	1832	4	US-09-568-472-4	Sequence 4, Appl
C 433	7	0.5	1163	5	PCT-US96-01311-44	Sequence 44, Appl	506	7	0.5	1832	4	US-09-567-899-4	Sequence 4, Appl
434	7	0.5	1193	4	US-09-227-725A-4	Sequence 4, Appl	507	7	0.5	1865	1	US-08-588-985-2	Sequence 2, Appl
435	7	0.5	1214	2	US-08-231-193A-54	Sequence 54, Appl	508	7	0.5	1865	1	US-08-971-988-2	Sequence 2, Appl
436	7	0.5	1214	3	US-08-486-273A-54	Sequence 54, Appl	509	7	0.5	2555	3	US-09-058-489-36	Sequence 36, Appl
437	7	0.5	1214	3	US-08-480-474-54	Sequence 54, Appl	510	7	0.5	2627	2	US-08-751-189-3	Sequence 3, Appl
438	7	0.5	1214	3	US-08-940-086A-54	Sequence 54, Appl	511	7	0.5	2627	2	US-09-060-836-3	Sequence 3, Appl
439	7	0.5	1214	4	US-08-940-035A-54	Sequence 54, Appl	512	7	0.5	2627	4	US-09-184-445-3	Sequence 3, Appl
440	7	0.5	1214	4	US-08-935-105A-54	Sequence 54, Appl	513	7	0.5	3066	4	US-08-952-127-12	Sequence 12, Appl
441	7	0.5	1214	4	US-09-648-797-54	Sequence 54, Appl	514	7	0.5	3177	2	US-08-477-451-4	Sequence 4, Appl
442	7	0.5	1219	2	US-08-231-193A-50	Sequence 50, Appl	515	7	0.5	15281	8	US-08-471-119A-2	Sequence 2, Appl
443	7	0.5	1219	3	US-08-486-273A-50	Sequence 50, Appl	516	6	0.5	8	1	US-08-076-092-64	Sequence 64, Appl
444	7	0.5	1219	3	US-08-480-474-50	Sequence 50, Appl	517	6	0.5	8	1	US-08-704-170-35	Sequence 35, Appl
445	7	0.5	1219	3	US-08-940-086A-50	Sequence 50, Appl	518	6	0.5	8	2	US-08-730-486-64	Sequence 64, Appl
446	7	0.5	1219	4	US-08-940-035A-50	Sequence 50, Appl	519	6	0.5	8	4	US-09-314-268-41	Sequence 41, Appl
447	7	0.5	1219	4	US-08-935-105A-50	Sequence 50, Appl	520	6	0.5	8	4	US-09-314-268-42	Sequence 42, Appl
448	7	0.5	1219	4	US-09-648-797-50	Sequence 50, Appl	521	6	0.5	8	4	US-09-314-268-43	Sequence 43, Appl
449	7	0.5	1226	1	US-08-280-443-2	Sequence 2, Appl	522	6	0.5	8	5	PCT-US94-02631-35	Sequence 35, Appl
450	7	0.5	1226	1	US-08-457-459-2	Sequence 2, Appl	523	6	0.5	9	1	US-08-159-340A-20	Sequence 20, Appl

524	6	0.5	9	1	US-08-076-092-62	Sequence 62, Appl	597	6	0.5	15	2	US-08-475-352-9	Sequence 9, Appl1
525	6	0.5	9	1	US-08-076-092-65	Sequence 65, Appl	598	6	0.5	15	4	US-08-602-999A-326	Sequence 326, App
526	6	0.5	9	1	US-08-421-702A-14	Sequence 14, Appl	599	6	0.5	15	4	US-08-602-999A-327	Sequence 327, App
527	6	0.5	9	1	US-08-303-052A-14	Sequence 14, Appl	600	6	0.5	15	4	US-08-602-999A-328	Sequence 328, App
528	6	0.5	9	1	US-08-421-698A-14	Sequence 14, Appl	601	6	0.5	15	4	US-08-602-999A-329	Sequence 329, App
529	6	0.5	9	1	US-08-421-697A-14	Sequence 14, Appl	602	6	0.5	15	4	US-08-602-999A-351	Sequence 351, App
530	6	0.5	9	1	US-08-421-698A-14	Sequence 14, Appl	603	6	0.5	15	4	US-08-468-337-19	Sequence 19, Appl
531	6	0.5	9	2	US-08-421-695A-14	Sequence 14, Appl	604	6	0.5	15	4	US-09-019-346A-12	Sequence 12, Appl
532	6	0.5	9	2	US-08-340-283-72	Sequence 72, Appl	605	6	0.5	15	4	US-09-500-124-326	Sequence 326, App
533	6	0.5	9	2	US-08-340-283-85	Sequence 85, Appl	606	6	0.5	15	4	US-09-500-124-327	Sequence 327, App
534	6	0.5	9	2	US-08-730-486-62	Sequence 62, Appl	607	6	0.5	15	4	US-09-500-124-328	Sequence 328, App
535	6	0.5	9	2	US-08-730-486-65	Sequence 65, Appl	608	6	0.5	15	4	US-09-500-124-329	Sequence 329, App
536	6	0.5	9	2	US-08-990-379-18	Sequence 18, Appl	609	6	0.5	15	4	US-09-500-124-351	Sequence 351, App
537	6	0.5	9	5	PCT-US95-04741-14	Sequence 14, Appl	610	6	0.5	15	5	PCT-US93-06751-37	Sequence 37, Appl
538	6	0.5	10	1	US-08-076-092-60	Sequence 60, Appl	611	6	0.5	15	5	PCT-US94-05796-9	Sequence 9, Appl1
539	6	0.5	10	1	US-08-076-092-63	Sequence 63, Appl	612	6	0.5	16	1	US-08-218-025A-12	Sequence 12, Appl
540	6	0.5	10	1	US-08-440-504A-10	Sequence 10, Appl	613	6	0.5	16	1	US-07-942-245-43	Sequence 43, Appl
541	6	0.5	10	2	US-08-716-317-32	Sequence 32, Appl	614	6	0.5	16	1	US-07-942-245-44	Sequence 44, Appl
542	6	0.5	10	2	US-08-747-137-80	Sequence 80, Appl	615	6	0.5	16	1	US-07-942-245-48	Sequence 48, Appl
543	6	0.5	10	2	US-08-730-486-60	Sequence 60, Appl	616	6	0.5	16	1	US-07-942-245-50	Sequence 50, Appl
544	6	0.5	10	2	US-08-730-486-63	Sequence 63, Appl	617	6	0.5	16	2	US-08-637-759B-465	Sequence 465, App
545	6	0.5	10	3	US-08-526-136-18	Sequence 18, Appl	618	6	0.5	16	3	US-08-871-355A-465	Sequence 465, App
546	6	0.5	10	4	US-09-319-730-8	Sequence 8, Appl1	619	6	0.5	16	3	US-09-037-190-27	Sequence 27, Appl
547	6	0.5	10	4	US-09-319-730-9	Sequence 9, Appl1	620	6	0.5	16	3	US-09-037-192-27	Sequence 27, Appl
548	6	0.5	10	5	PCT-US95-03236-17	Sequence 17, Appl	621	6	0.5	16	3	US-09-037-143-27	Sequence 27, Appl
549	6	0.5	11	1	US-08-076-092-5	Sequence 5, Appl1	622	6	0.5	16	4	US-09-049-691-27	Sequence 27, Appl
550	6	0.5	11	1	US-08-076-092-59	Sequence 59, Appl	623	6	0.5	16	4	US-08-260-174-27	Sequence 27, Appl
551	6	0.5	11	1	US-08-076-092-61	Sequence 61, Appl	624	6	0.5	16	4	US-09-338-128A-27	Sequence 27, Appl
552	6	0.5	11	2	US-08-705-660-1	Sequence 1, Appl1	c 625	6	0.5	16	4	US-09-201-945-465	Sequence 465, App
553	6	0.5	11	2	US-08-730-486-5	Sequence 5, Appl1	626	6	0.5	16	4	US-09-232-346-27	Sequence 27, Appl
554	6	0.5	11	2	US-08-730-486-59	Sequence 59, Appl	627	6	0.5	16	4	US-09-037-192-27	Sequence 27, Appl
555	6	0.5	11	2	US-08-730-486-61	Sequence 61, Appl	628	6	0.5	16	4	US-09-288-339-24	Sequence 24, Appl
556	6	0.5	11	3	US-08-989-045-1	Sequence 1, Appl1	629	6	0.5	16	6	5378805-3	Patent No. 5378805
557	6	0.5	11	4	US-08-817-177-4	Sequence 4, Appl1	630	6	0.5	17	1	US-08-370-567-9	Sequence 9, Appl1
558	6	0.5	11	4	US-09-227-357-596	Sequence 596, Appl	631	6	0.5	17	1	US-08-438-759-9	Sequence 9, Appl1
559	6	0.5	11	4	US-09-288-339-14	Sequence 14, Appl	632	6	0.5	17	1	US-08-538-911-21	Sequence 21, Appl
560	6	0.5	11	5	PCT-US95-12686-4	Sequence 4, Appl1	633	6	0.5	17	4	US-09-308-022-3	Sequence 3, Appl1
561	6	0.5	12	1	US-07-939-501A-5	Sequence 5, Appl1	634	6	0.5	17	5	US-09-025-769B-229	Sequence 229, App
562	6	0.5	12	1	US-08-076-092-58	Sequence 58, Appl	635	6	0.5	17	5	PCT-US94-05591-21	Sequence 21, Appl
563	6	0.5	12	1	US-08-240-514-24	Sequence 24, Appl	636	6	0.5	17	5	PCT-US94-05684-9	Sequence 9, Appl1
564	6	0.5	12	1	US-08-548-540-129	Sequence 129, App	637	6	0.5	18	1	US-08-488-212A-30	Sequence 30, Appl
565	6	0.5	12	2	US-08-612-302A-24	Sequence 24, Appl	638	6	0.5	18	1	US-08-207-169A-7	Sequence 7, Appl1
566	6	0.5	12	2	US-08-406-330-24	Sequence 24, Appl	639	6	0.5	18	1	US-08-338-127-2	Sequence 2, Appl1
567	6	0.5	12	2	US-08-392-973A-2	Sequence 2, Appl1	640	6	0.5	18	2	US-08-320-306-30	Sequence 30, Appl
568	6	0.5	12	2	US-08-556-597-24	Sequence 24, Appl	641	6	0.5	18	2	US-08-488-209B-30	Sequence 30, Appl
569	6	0.5	12	2	US-08-730-486-58	Sequence 58, Appl	642	6	0.5	18	2	US-08-408-011-30	Sequence 30, Appl
570	6	0.5	12	4	US-09-258-754-89	Sequence 89, Appl	c 643	6	0.5	18	2	US-09-017-205-47	Sequence 47, Appl
571	6	0.5	12	4	US-08-448-398-9	Sequence 9, Appl1	644	6	0.5	18	2	US-09-017-205-48	Sequence 48, Appl
572	6	0.5	12	4	US-09-042-107-89	Sequence 89, Appl	645	6	0.5	19	1	US-07-596-081A-28	Sequence 28, Appl
573	6	0.5	12	5	PCT-US93-12679-26	Sequence 26, Appl	646	6	0.5	19	1	US-08-019-073-30	Sequence 30, Appl
574	6	0.5	12	5	PCT-US96-09809-129	Sequence 129, App	647	6	0.5	19	2	US-08-764-640-158	Sequence 158, App
575	6	0.5	13	3	US-09-037-190-23	Sequence 23, Appl	648	6	0.5	19	2	US-08-483-636-8	Sequence 8, Appl1
576	6	0.5	13	3	US-09-037-192-23	Sequence 23, Appl	649	6	0.5	19	2	US-08-483-632-8	Sequence 8, Appl1
577	6	0.5	13	4	US-09-037-143-23	Sequence 23, Appl	650	6	0.5	19	3	US-09-100-414B-17	Sequence 17, Appl
578	6	0.5	13	4	US-09-049-691-23	Sequence 23, Appl	651	6	0.5	19	3	US-08-973-225-158	Sequence 158, App
579	6	0.5	13	4	US-08-602-999A-83	Sequence 83, Appl	652	6	0.5	19	3	US-09-244-298A-158	Sequence 158, App
580	6	0.5	13	4	US-08-260-174-23	Sequence 23, Appl	653	6	0.5	19	4	US-09-303-323-17	Sequence 17, Appl
581	6	0.5	13	4	US-08-278-865-83	Sequence 83, Appl	654	6	0.5	19	4	US-08-928-213B-132	Sequence 132, App
582	6	0.5	13	4	US-09-319-730-2	Sequence 2, Appl1	c 655	6	0.5	19	4	US-09-516-704-158	Sequence 158, App
583	6	0.5	13	4	US-09-338-128A-23	Sequence 23, Appl	656	6	0.5	19	4	US-09-441-502B-11	Sequence 11, Appl
584	6	0.5	13	4	US-09-232-346-23	Sequence 23, Appl	657	6	0.5	19	4	US-09-549-090-158	Sequence 158, App
585	6	0.5	13	4	US-09-037-192-23	Sequence 23, Appl	658	6	0.5	19	5	PCT-US94-01768-30	Sequence 30, Appl
586	6	0.5	13	4	US-08-182-967-15	Sequence 15, Appl	659	6	0.5	20	1	US-07-700-526-15	Sequence 15, Appl
587	6	0.5	13	4	US-09-500-124-83	Sequence 83, Appl	660	6	0.5	20	1	US-07-678-974D-43	Sequence 43, Appl
588	6	0.5	14	1	US-08-320-373-9	Sequence 9, Appl1	661	6	0.5	20	1	US-07-678-974D-53	Sequence 53, Appl
589	6	0.5	14	2	US-08-442-461D-4	Sequence 4, Appl1	662	6	0.5	20	1	US-08-338-127-1	Sequence 1, Appl1
590	6	0.5	14	4	US-09-426-680-4	Sequence 4, Appl1	663	6	0.5	20	1	US-08-222-851-40	Sequence 40, Appl
591	6	0.5	14	4	US-09-608-285A-44	Sequence 44, Appl	664	6	0.5	20	1	US-08-732-751-10	Sequence 10, Appl
592	6	0.5	15	1	US-07-978-895-9	Sequence 9, Appl1	665	6	0.5	20	2	US-08-934-915-10	Sequence 10, Appl
593	6	0.5	15	1	US-08-067-387-9	Sequence 9, Appl1	666	6	0.5	20	2	US-08-934-915-17	Sequence 17, Appl
594	6	0.5	15	1	US-08-320-373-50	Sequence 50, Appl	667	6	0.5	20	2	US-08-934-915-40	Sequence 40, Appl
595	6	0.5	15	1	US-08-237-418-19	Sequence 19, Appl	668	6	0.5	20	2	US-08-934-915-141	Sequence 141, App
596	6	0.5	15	2	US-08-473-119-9	Sequence 9, Appl1	669	6	0.5	20	2	US-08-934-915-170	Sequence 170, App

c 670	6	0.5	20	2	US-08-945-168-49	Sequence 49, Appl	c 743	6	0.5	27	3	US-08-834-306-46	Sequence 46, Appl
c 671	6	0.5	20	2	US-08-945-168-59	Sequence 59, Appl	c 744	6	0.5	27	4	US-08-993-674A-46	Sequence 46, Appl
c 672	6	0.5	20	2	US-09-010-928B-7	Sequence 7, Appl	c 745	6	0.5	27	4	US-09-406-781-14	Sequence 14, Appl
c 673	6	0.5	20	3	US-08-460-576-15	Sequence 15, Appl	c 746	6	0.5	27	4	US-09-256-976-46	Sequence 46, Appl
c 674	6	0.5	20	4	US-08-602-999A-110	Sequence 110, App	c 747	6	0.5	28	1	US-08-485-692-13	Sequence 13, Appl
c 675	6	0.5	20	4	US-08-602-999A-116	Sequence 116, App	c 748	6	0.5	28	1	US-08-732-751-11	Sequence 11, Appl
c 676	6	0.5	20	4	US-08-928-213B-119	Sequence 119, App	c 749	6	0.5	28	1	US-08-419-519-13	Sequence 13, Appl
c 677	6	0.5	20	4	US-08-928-213B-123	Sequence 123, App	c 750	6	0.5	28	4	US-09-177-249-97	Sequence 97, Appl
c 678	6	0.5	20	4	US-08-928-213B-124	Sequence 124, App	c 751	6	0.5	28	4	US-08-918-288-1	Sequence 1, Appl
c 679	6	0.5	20	4	US-08-928-213B-125	Sequence 125, App	c 752	6	0.5	28	4	US-09-282-357-1	Sequence 1, Appl
c 680	6	0.5	20	4	US-08-928-213B-127	Sequence 127, App	c 753	6	0.5	28	4	US-09-604-871-3	Sequence 3, Appl
c 681	6	0.5	20	4	US-08-928-213B-128	Sequence 128, App	c 754	6	0.5	29	1	US-07-960-510-4	Sequence 4, Appl
c 682	6	0.5	20	4	US-08-928-213B-129	Sequence 129, App	c 755	6	0.5	30	1	US-08-732-751-7	Sequence 7, Appl
c 683	6	0.5	20	4	US-08-928-213B-130	Sequence 130, App	c 756	6	0.5	30	2	US-08-753-829A-14	Sequence 14, Appl
c 684	6	0.5	20	4	US-08-817-895-12	Sequence 12, Appl	c 757	6	0.5	30	2	US-08-716-317-16	Sequence 16, Appl
c 685	6	0.5	20	4	US-09-500-124-110	Sequence 110, App	c 758	6	0.5	30	2	US-08-934-915-42	Sequence 42, Appl
c 686	6	0.5	20	4	US-09-500-124-116	Sequence 116, App	c 759	6	0.5	30	2	US-08-934-915-190	Sequence 190, App
c 687	6	0.5	20	4	US-08-651-650-19	Sequence 19, Appl	c 760	6	0.5	30	2	US-08-934-915-193	Sequence 193, App
c 688	6	0.5	20	5	PCT-US92-03132-15	Sequence 15, Appl	c 761	6	0.5	30	2	US-08-031-538-50	Sequence 50, Appl
c 689	6	0.5	20	5	PCT-US94-12985-2	Sequence 2, Appl	c 762	6	0.5	30	4	US-09-037-179B-14	Sequence 14, Appl
c 690	6	0.5	20	5	PCT-US94-12985-8	Sequence 8, Appl	c 763	6	0.5	30	4	US-09-037-179B-17	Sequence 17, Appl
c 691	6	0.5	21	1	US-07-969-931-2	Sequence 2, Appl	c 764	6	0.5	31	1	US-07-596-081A-1	Sequence 1, Appl
c 692	6	0.5	21	1	US-07-855-417A-2	Sequence 2, Appl	c 765	6	0.5	31	1	US-07-596-081A-24	Sequence 24, Appl
c 693	6	0.5	21	2	US-08-894-339-8	Sequence 8, Appl	c 766	6	0.5	31	1	US-08-732-751-3	Sequence 3, Appl
c 694	6	0.5	21	2	US-08-454-236-9	Sequence 9, Appl	c 767	6	0.5	31	1	US-08-732-751-4	Sequence 4, Appl
c 695	6	0.5	21	2	US-09-092-770-16	Sequence 16, Appl	c 768	6	0.5	31	2	US-08-455-625-6	Sequence 6, Appl
c 696	6	0.5	21	4	US-09-222-851-16	Sequence 16, Appl	c 769	6	0.5	31	2	US-08-455-625-33	Sequence 33, Appl
c 697	6	0.5	21	4	US-09-306-044-8	Sequence 8, Appl	c 770	6	0.5	31	3	US-09-100-414B-55	Sequence 55, Appl
c 698	6	0.5	21	4	US-09-227-357-284	Sequence 284, App	c 771	6	0.5	31	4	US-08-455-685-6	Sequence 6, Appl
c 699	6	0.5	22	1	US-07-596-081A-7	Sequence 7, Appl	c 772	6	0.5	31	4	US-08-455-685-33	Sequence 33, Appl
c 700	6	0.5	22	1	US-08-103-445-15	Sequence 15, Appl	c 773	6	0.5	31	4	US-09-303-323-55	Sequence 55, Appl
c 701	6	0.5	22	1	US-08-461-690B-15	Sequence 15, Appl	c 774	6	0.5	31	4	US-08-060-988A-6	Sequence 6, Appl
c 702	6	0.5	22	1	US-08-241-853-37	Sequence 37, Appl	c 775	6	0.5	31	4	US-08-060-988A-33	Sequence 33, Appl
c 703	6	0.5	22	2	US-08-850-917-37	Sequence 37, Appl	c 776	6	0.5	31	5	PCT-US94-05142-6	Sequence 6, Appl
c 704	6	0.5	22	2	US-08-525-742-23	Sequence 23, Appl	c 777	6	0.5	31	5	PCT-US94-05142-33	Sequence 33, Appl
c 705	6	0.5	22	2	US-08-194-881B-18	Sequence 18, Appl	c 778	6	0.5	32	1	US-08-152-488-9	Sequence 9, Appl
c 706	6	0.5	22	3	US-08-783-853A-39	Sequence 39, Appl	c 779	6	0.5	32	1	US-08-303-025-9	Sequence 9, Appl
c 707	6	0.5	22	4	US-09-406-781-41	Sequence 41, Appl	c 780	6	0.5	32	1	US-08-677-304-9	Sequence 9, Appl
c 708	6	0.5	22	4	US-09-344-050-39	Sequence 39, Appl	c 781	6	0.5	32	2	US-08-716-317-21	Sequence 21, Appl
c 709	6	0.5	22	4	US-09-149-476-320	Sequence 320, App	c 782	6	0.5	32	2	US-08-436-703B-14	Sequence 14, Appl
c 710	6	0.5	22	4	US-09-149-476-506	Sequence 506, App	c 783	6	0.5	32	2	US-09-252-586-15	Sequence 15, Appl
c 711	6	0.5	22	4	US-09-149-476-507	Sequence 507, App	c 784	6	0.5	33	1	US-07-596-081A-6	Sequence 6, Appl
c 712	6	0.5	22	6	5204326-76	Patent No. 5204326	c 785	6	0.5	33	2	US-08-146-028-32	Sequence 32, Appl
c 713	6	0.5	23	2	US-08-460-890A-56	Sequence 56, Appl	c 786	6	0.5	33	2	US-08-146-028-171	Sequence 171, App
c 714	6	0.5	23	3	US-08-783-853A-24	Sequence 24, Appl	c 787	6	0.5	33	4	US-08-723-425A-32	Sequence 32, Appl
c 715	6	0.5	23	3	US-08-783-853A-76	Sequence 76, Appl	c 788	6	0.5	33	4	US-08-723-425A-171	Sequence 171, App
c 716	6	0.5	23	3	US-08-167-641C-56	Sequence 56, Appl	c 789	6	0.5	33	4	US-09-112-206-32	Sequence 32, Appl
c 717	6	0.5	23	4	US-08-460-971A-56	Sequence 56, Appl	c 790	6	0.5	33	4	US-09-112-206-171	Sequence 171, App
c 718	6	0.5	23	4	US-08-462-040-56	Sequence 56, Appl	c 791	6	0.5	33	4	US-08-857-046A-31	Sequence 31, Appl
c 719	6	0.5	23	4	US-09-344-050-24	Sequence 24, Appl	c 792	6	0.5	33	4	US-09-149-476-447	Sequence 447, App
c 720	6	0.5	23	4	US-09-344-050-76	Sequence 76, Appl	c 793	6	0.5	33	4	US-09-149-476-447	Sequence 447, App
c 721	6	0.5	23	4	US-08-835-281-5	Sequence 5, Appl	c 794	6	0.5	34	1	US-09-570-921-126	Sequence 126, App
c 722	6	0.5	23	4	US-08-835-281-6	Sequence 6, Appl	c 795	6	0.5	34	1	US-07-939-501A-21	Sequence 21, Appl
c 723	6	0.5	23	6	5451527-3	Patent No. 5451527	c 796	6	0.5	34	1	US-08-700-749A-3	Sequence 3, Appl
c 724	6	0.5	24	1	US-09-732-751-5	Sequence 5, Appl	c 797	6	0.5	34	2	US-08-716-317-26	Sequence 26, Appl
c 725	6	0.5	24	4	US-09-217-306B-24	Sequence 24, Appl	c 798	6	0.5	34	2	US-08-454-236-6	Sequence 6, Appl
c 726	6	0.5	25	2	US-08-942-423-26	Sequence 26, Appl	c 799	6	0.5	34	3	US-09-020-684-3	Sequence 3, Appl
c 727	6	0.5	25	4	US-09-227-357-320	Sequence 320, App	c 800	6	0.5	34	3	US-09-020-685-3	Sequence 3, Appl
c 728	6	0.5	25	4	US-08-826-134-16	Sequence 16, Appl	c 801	6	0.5	34	3	US-09-020-685-3	Sequence 3, Appl
c 729	6	0.5	25	4	US-08-826-134-27	Sequence 27, Appl	c 802	6	0.5	34	4	US-09-020-683-3	Sequence 3, Appl
c 730	6	0.5	25	5	PCT-US93-05325-10	Sequence 10, Appl	c 803	6	0.5	34	4	US-09-227-357-321	Sequence 321, App
c 731	6	0.5	26	1	US-07-942-245-378	Sequence 378, App	c 804	6	0.5	34	6	5451527-4	Patent No. 5451527
c 732	6	0.5	26	2	US-08-620-151-83	Sequence 83, Appl	c 805	6	0.5	35	2	US-08-867-087B-60	Sequence 60, Appl
c 733	6	0.5	26	3	US-08-755-587-217	Sequence 217, App	c 806	6	0.5	35	2	US-08-867-087B-61	Sequence 61, Appl
c 734	6	0.5	26	4	US-08-602-999A-51	Sequence 51, Appl	c 807	6	0.5	35	3	US-09-053-197A-63	Sequence 63, Appl
c 735	6	0.5	26	4	US-09-500-124-51	Sequence 51, Appl	c 808	6	0.5	35	3	US-08-905-223-331	Sequence 331, App
c 736	6	0.5	27	1	US-08-786-748A-83	Sequence 83, Appl	c 809	6	0.5	35	4	US-09-085-761A-68	Sequence 68, Appl
c 737	6	0.5	27	1	US-08-786-748A-84	Sequence 84, Appl	c 810	6	0.5	35	4	US-09-314-268-117	Sequence 117, App
c 738	6	0.5	27	1	US-08-732-751-12	Sequence 12, Appl	c 811	6	0.5	36	1	US-08-237-418-20	Sequence 20, Appl
c 739	6	0.5	27	1	US-08-557-309B-46	Sequence 46, Appl	c 812	6	0.5	36	4	US-08-468-337-20	Sequence 20, Appl
c 740	6	0.5	27	2	US-08-932-682-83	Sequence 83, Appl	c 813	6	0.5	36	4	US-09-215-212-8	Sequence 8, Appl
c 741	6	0.5	27	2	US-08-932-682-84	Sequence 84, Appl	c 814	6	0.5	36	4	US-09-092-315-19	Sequence 19, Appl
c 742	6	0.5	27	3	US-08-654-623-23	Sequence 23, Appl	c 815	6	0.5	36	4	US-08-469-260A-414	Sequence 414, App

c 816	6	0.5	36	6	5451527-5	Patent No. 5451527	889	6	0.5	45	4	US-09-314-268-102	Sequence 102, App
c 817	6	0.5	36	6	5451527-6	Patent No. 5451527	c 890	6	0.5	45	4	US-09-314-268-102	Sequence 102, App
c 818	6	0.5	37	2	US-08-709-924-25	Sequence 25, Appl	c 891	6	0.5	45	4	US-09-314-268-103	Sequence 103, App
c 819	6	0.5	37	2	US-08-709-925-25	Sequence 25, Appl	c 892	6	0.5	45	6	5451527-1	Patent No. 5451527
c 820	6	0.5	37	4	US-08-709-948-25	Sequence 25, Appl	c 893	6	0.5	46	4	US-09-004-406C-14	Sequence 14, Appl
c 821	6	0.5	37	4	US-09-149-476-694	Sequence 694, App	c 894	6	0.5	46	4	US-08-469-260A-256	Sequence 256, App
c 822	6	0.5	38	1	US-08-176-500-25	Sequence 25, Appl	c 895	6	0.5	47	2	US-09-227-357-637	Sequence 637, App
c 823	6	0.5	38	1	US-08-176-500-131	Sequence 131, App	c 896	6	0.5	48	2	US-08-637-759B-199	Sequence 199, App
c 824	6	0.5	38	1	US-08-471-052A-25	Sequence 25, Appl	c 897	6	0.5	48	3	US-08-871-355A-199	Sequence 199, App
c 825	6	0.5	38	1	US-08-471-052A-131	Sequence 131, App	c 898	6	0.5	48	4	US-09-201-945-199	Sequence 199, App
c 826	6	0.5	38	1	US-08-189-331-25	Sequence 25, Appl	c 899	6	0.5	49	2	US-08-765-179B-7	Sequence 7, Appl
c 827	6	0.5	38	1	US-08-189-331-131	Sequence 131, App	c 900	6	0.5	49	4	US-09-227-357-595	Sequence 595, App
c 828	6	0.5	38	2	US-08-471-939-25	Sequence 25, Appl	c 901	6	0.5	49	4	US-09-065-383-31	Sequence 31, Appl
c 829	6	0.5	38	2	US-08-471-939-131	Sequence 131, App	c 902	6	0.5	50	4	US-09-314-268-159	Sequence 159, App
c 830	6	0.5	38	2	US-08-471-939-131	Sequence 25, Appl	c 903	6	0.5	51	1	US-08-445-909A-18	Sequence 18, Appl
c 831	6	0.5	38	2	US-08-471-800-131	Sequence 131, App	c 904	6	0.5	51	4	US-09-230-637-46	Sequence 46, Appl
c 832	6	0.5	38	2	US-08-488-161-14	Sequence 14, Appl	c 905	6	0.5	52	1	US-08-346-849-8	Sequence 8, Appl
c 833	6	0.5	38	2	US-08-471-068-25	Sequence 25, Appl	c 906	6	0.5	52	2	US-08-293-284A-8	Sequence 8, Appl
c 834	6	0.5	38	2	US-08-471-068-131	Sequence 131, App	c 907	6	0.5	52	3	US-08-917-299-32	Sequence 32, Appl
c 835	6	0.5	38	2	US-08-726-306A-88	Sequence 88, Appl	c 908	6	0.5	52	4	US-09-422-662-32	Sequence 32, Appl
c 836	6	0.5	38	3	US-09-273-685-14	Sequence 14, Appl	c 909	6	0.5	52	4	US-09-153-599A-4	Sequence 4, Appl
c 837	6	0.5	38	5	PCR-US95-11934-14	Sequence 14, Appl	c 910	6	0.5	53	2	US-08-852-401-5	Sequence 5, Appl
c 838	6	0.5	39	2	US-08-657-392-29	Sequence 29, Appl	c 911	6	0.5	53	4	US-09-149-476-448	Sequence 448, App
c 839	6	0.5	39	2	US-08-488-161-33	Sequence 33, Appl	c 912	6	0.5	54	4	US-08-905-223-427	Sequence 427, App
c 840	6	0.5	39	3	US-09-273-685-33	Sequence 33, Appl	c 913	6	0.5	54	4	US-09-306-446C-13	Sequence 13, Appl
c 841	6	0.5	39	4	US-09-314-268-155	Sequence 155, App	c 914	6	0.5	55	3	US-08-814-836-7	Sequence 7, Appl
c 842	6	0.5	39	4	US-09-314-268-156	Sequence 156, App	c 915	6	0.5	55	4	US-09-314-268-147	Sequence 147, App
c 843	6	0.5	39	4	US-09-384-302A-7	Sequence 7, Appl	c 916	6	0.5	56	4	US-09-227-357-577	Sequence 577, App
c 844	6	0.5	39	5	PCR-US94-02539-29	Sequence 29, Appl	c 917	6	0.5	57	1	US-08-137-614A-17	Sequence 17, Appl
c 845	6	0.5	39	5	PCR-US95-11934-33	Sequence 33, Appl	c 918	6	0.5	57	3	US-08-814-836-17	Sequence 17, Appl
c 846	6	0.5	39	6	5451527-8	Patent No. 5451527	c 919	6	0.5	58	1	US-08-406-809-3	Sequence 3, Appl
c 847	6	0.5	39	6	5451527-12	Patent No. 5451527	c 920	6	0.5	58	1	US-08-406-809-4	Sequence 4, Appl
c 848	6	0.5	40	1	US-08-188-228-18	Sequence 18, Appl	c 921	6	0.5	58	1	US-08-401-908-31	Sequence 31, Appl
c 849	6	0.5	40	1	US-08-332-643-18	Sequence 18, Appl	c 922	6	0.5	58	4	US-09-461-697-173	Sequence 173, App
c 850	6	0.5	40	1	US-08-332-638-18	Sequence 18, Appl	c 923	6	0.5	58	4	US-09-319-730-17	Sequence 17, Appl
c 851	6	0.5	40	2	US-08-194-981B-14	Sequence 14, Appl	c 924	6	0.5	59	1	US-08-287-959-21	Sequence 21, Appl
c 852	6	0.5	40	2	US-08-194-981B-15	Sequence 15, Appl	c 925	6	0.5	59	2	US-08-716-317-2	Sequence 2, Appl
c 853	6	0.5	40	3	US-09-037-524-4	Sequence 4, Appl	c 926	6	0.5	59	3	US-08-657-983A-4	Sequence 4, Appl
c 854	6	0.5	40	4	US-08-988-856B-7	Sequence 7, Appl	c 927	6	0.5	59	4	US-09-036-549A-4	Sequence 4, Appl
c 855	6	0.5	40	4	US-08-908-371B-10	Sequence 10, Appl	c 928	6	0.5	59	4	US-08-924-629C-30	Sequence 30, Appl
c 856	6	0.5	41	1	US-07-709-091-3	Sequence 3, Appl	c 929	6	0.5	59	4	US-09-082-358B-50	Sequence 50, Appl
c 857	6	0.5	41	1	US-07-715-752A-3	Sequence 3, Appl	c 930	6	0.5	60	2	US-08-117-952-788	Sequence 788, App
c 858	6	0.5	41	1	US-08-104-862-3	Sequence 3, Appl	c 931	6	0.5	60	2	US-08-607-412-2	Sequence 2, Appl
c 859	6	0.5	41	1	US-08-162-178-3	Sequence 3, Appl	c 932	6	0.5	60	4	US-09-227-357-307	Sequence 307, App
c 860	6	0.5	41	1	US-08-122-520C-2	Sequence 2, Appl	c 933	6	0.5	60	4	US-09-227-357-317	Sequence 317, App
c 861	6	0.5	41	1	US-08-865-773-4	Sequence 4, Appl	c 934	6	0.5	60	4	US-09-227-357-518	Sequence 518, App
c 862	6	0.5	41	2	US-08-194-981B-17	Sequence 17, Appl	c 935	6	0.5	60	4	US-08-924-629C-31	Sequence 31, Appl
c 863	6	0.5	41	4	US-09-174-465D-8	Sequence 8, Appl	c 936	6	0.5	60	4	US-09-336-536-36	Sequence 36, Appl
c 864	6	0.5	41	4	US-09-424-127-4	Sequence 4, Appl	c 937	6	0.5	61	4	US-09-319-730-15	Sequence 15, Appl
c 865	6	0.5	41	4	US-09-314-268-104	Sequence 104, App	c 938	6	0.5	61	4	US-08-858-207A-538	Sequence 538, App
c 866	6	0.5	41	4	US-08-936-165A-290	Sequence 290, App	c 939	6	0.5	63	6	5164482-9	Patent No. 5164482
c 867	6	0.5	41	4	US-09-599-564A-8	Sequence 8, Appl	c 940	6	0.5	64	1	US-08-470-202-39	Sequence 39, Appl
c 868	6	0.5	41	4	US-09-201-227A-27	Sequence 27, Appl	c 941	6	0.5	64	1	US-08-471-770-39	Sequence 39, Appl
c 869	6	0.5	41	5	PCR-US92-05101-3	Sequence 3, Appl	c 942	6	0.5	64	2	US-08-292-870-1	Sequence 1, Appl
c 870	6	0.5	41	6	5451527-9	Patent No. 5451527	c 943	6	0.5	64	2	US-08-468-059-39	Sequence 39, Appl
c 871	6	0.5	41	6	5451527-10	Patent No. 5451527	c 944	6	0.5	64	4	US-09-109-916-39	Sequence 39, Appl
c 872	6	0.5	41	6	5451527-11	Patent No. 5451527	c 945	6	0.5	64	4	US-09-134-001C-4846	Sequence 4846, Ap
c 873	6	0.5	41	6	5451527-13	Patent No. 5451527	c 946	6	0.5	64	4	US-08-469-260A-384	Sequence 384, App
c 874	6	0.5	41	6	5451527-14	Patent No. 5451527	c 947	6	0.5	66	4	US-09-227-357-160	Sequence 160, App
c 875	6	0.5	42	1	US-07-651-710A-23	Sequence 23, App	c 948	6	0.5	67	2	US-08-273-146-67	Sequence 67, Appl
c 876	6	0.5	42	3	US-08-924-330A-3	Sequence 3, Appl	c 949	6	0.5	67	4	US-09-331-254-2	Sequence 2, Appl
c 877	6	0.5	42	4	US-09-138-721-3	Sequence 3, Appl	c 950	6	0.5	68	1	US-07-756-250-10	Sequence 10, Appl
c 878	6	0.5	43	4	US-09-156-580-9	Sequence 9, Appl	c 951	6	0.5	68	2	US-08-511-485-16	Sequence 16, Appl
c 879	6	0.5	43	4	US-08-905-223-293	Sequence 293, App	c 952	6	0.5	68	4	US-09-220-528-50	Sequence 50, Appl
c 880	6	0.5	43	4	US-09-156-579C-11	Sequence 11, Appl	c 953	6	0.5	69	2	US-08-997-080-184	Sequence 184, App
c 881	6	0.5	43	4	US-08-908-371B-12	Sequence 12, Appl	c 954	6	0.5	69	2	US-08-997-362-184	Sequence 184, App
c 882	6	0.5	44	2	US-08-484-434C-24	Sequence 24, Appl	c 955	6	0.5	69	4	US-09-095-855-184	Sequence 184, App
c 883	6	0.5	44	3	US-08-707-399B-18	Sequence 18, Appl	c 956	6	0.5	69	4	US-09-324-542-184	Sequence 184, App
c 884	6	0.5	45	2	US-08-530-290-21	Sequence 21, Appl	c 957	6	0.5	69	4	US-09-134-001C-3959	Sequence 3959, Ap
c 885	6	0.5	45	3	US-09-100-600A-56	Sequence 56, Appl	c 958	6	0.5	69	4	US-09-205-426-184	Sequence 184, App
c 886	6	0.5	45	3	US-08-236-886-2	Sequence 2, Appl	c 959	6	0.5	70	4	US-09-367-953B-68	Sequence 68, Appl
c 887	6	0.5	45	4	US-08-857-076-62	Sequence 62, Appl	c 960	6	0.5	71	2	US-08-972-008-4	Sequence 4, Appl
c 888	6	0.5	45	4	US-08-857-076-92	Sequence 92, Appl	c 961	6	0.5	71	4	US-09-134-001C-4418	Sequence 4418, Ap

```
c 962      6      0.5      71      4      US-08-267-409-4      Sequence 4, Appli
c 963      6      0.5      72      1      US-08-439-818A-21      Sequence 21, Appl
c 964      6      0.5      72      1      US-08-542-363-17      Sequence 17, Appl
c 965      6      0.5      72      2      US-08-751-965-21      Sequence 21, Appl
c 966      6      0.5      72      2      US-08-738-975-21      Sequence 21, Appl
c 967      6      0.5      72      2      US-08-728-626-21      Sequence 21, Appl
c 968      6      0.5      72      2      US-08-530-290-19      Sequence 19, Appl
c 969      6      0.5      72      3      US-08-808-599A-21      Sequence 21, Appl
c 970      6      0.5      72      4      US-09-100-089-17      Sequence 17, Appl
c 971      6      0.5      72      4      US-08-905-223-314      Sequence 314, App
c 972      6      0.5      72      4      US-09-670-827-17      Sequence 17, Appl
c 973      6      0.5      73      4      US-09-509-902A-8       Sequence 8, Appli
c 974      6      0.5      73      4      US-08-469-260A-510     Sequence 510, App
c 975      6      0.5      74      2      US-08-343-443B-11      Sequence 11, Appl
c 976      6      0.5      74      4      US-09-337-227C-30      Sequence 30, Appl
c 977      6      0.5      75      1      US-07-881-075-12      Sequence 12, Appl
c 978      6      0.5      75      1      US-08-120-827-12      Sequence 12, Appl
c 979      6      0.5      75      1      US-08-478-675-12      Sequence 12, Appl
c 980      6      0.5      75      4      US-09-134-001C-3025   Sequence 3025, Ap
c 981      6      0.5      75      4      US-09-134-001C-5076   Sequence 5076, Ap
c 982      6      0.5      76      4      US-09-227-357-456     Sequence 456, App
c 983      6      0.5      76      4      US-09-281-766-12      Sequence 12, Appl
c 984      6      0.5      77      2      US-08-476-866-21      Sequence 21, Appl
c 985      6      0.5      78      2      US-08-469-412A-11      Sequence 11, Appl
c 986      6      0.5      78      4      US-09-021-715-11      Sequence 11, Appl
c 987      6      0.5      79      3      US-09-143-776-6       Sequence 6, Appli
c 988      6      0.5      80      2      US-08-332-562A-84      Sequence 84, Appl
c 989      6      0.5      81      4      US-09-249-542-2       Sequence 2, Appli
c 990      6      0.5      83      4      US-08-936-165A-410     Sequence 410, App
c 991      6      0.5      85      4      US-08-711-417C-190     Sequence 190, App
c 992      6      0.5      85      4      US-08-263-933-6       Sequence 6, Appli
c 993      6      0.5      85      4      US-09-263-933-13      Sequence 13, Appl
c 994      6      0.5      85      4      US-09-263-933-20      Sequence 20, Appl
c 995      6      0.5      86      4      US-08-858-207A-471     Sequence 471, App
c 996      6      0.5      87      4      US-08-711-417C-192     Sequence 192, App
c 997      6      0.5      87      4      US-08-711-417C-193     Sequence 193, App
c 998      6      0.5      87      4      US-08-220-528-115     Sequence 115, App
c 999      6      0.5      87      4      US-09-152-060-89      Sequence 89, Appl
c1000     6      0.5      88      1      US-08-469-202-20      Sequence 20, Appl
```

ALIGNMENTS

RESULT 1

```
us-09-149-476-477
; Sequence 477, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
```

```
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
```

/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,630
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,878
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,662
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,872
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,882
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,637
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,903
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,888
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,879
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,880
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,894
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,911
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,636
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,631
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,845
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,892
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/057,761
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/047,595
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,599
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,588
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,585
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,586
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,590
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,594
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,589
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,593
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,614
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,578
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,576
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/047,501
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,670
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/056,632
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,664
/ EARLIER FILING DATE: 1997-08-22

/ EARLIER APPLICATION NUMBER: 60/056,876
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,881
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,909
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,875
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,862
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,887
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,908
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/048,964
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/057,650
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/056,884
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/057,669
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/049,610
/ EARLIER FILING DATE: 1997-06-13
/ EARLIER APPLICATION NUMBER: 60/061,060
/ EARLIER FILING DATE: 1997-10-02

Alignment Scores:
Pred. No.: 75.4 Length: 192
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 4 Gaps: 0

US-09-818-990b-1 (1-3963) x US-09-149-476-477 (1-192)

QY 174 ATCTGGCCTCGCCTCTTCAGC 151

Db 9 IleuAlaSerAlaSerPheSer 16

RESULT 2

US-09-500-569-2
/ Sequence 2, Application US/09500569
/ Patent No. 6329204
/ GENERAL INFORMATION:
/ APPLICANT: Cahoon, Rebecca E.
/ APPLICANT: Rafalski, Antoni
/ TITLE OF INVENTION: Plant Caffeic acid 3-O-Methyltransferase Homologs
/ FILE REFERENCE: BB1327 US NA
/ CURRENT APPLICATION NUMBER: US/09/500,569
/ CURRENT FILING DATE: 2000-02-09
/ EARLIER APPLICATION NUMBER: 60/119,587
/ EARLIER FILING DATE: 1999-February-10
/ NUMBER OF SEQ ID NOS: 28
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 2
/ LENGTH: 305
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (217)
US-09-500-569-2

Alignment Scores:
Pred. No.: 70.5 Length: 305
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-500-569-2 (1-305)

QY 2216 GGGCCACAGTTGCTGGGTGGTG 2193

Db 216 GlyAlaThrValAlaAlaVal 223

RESULT 3

US-08-440-845D-5
; Sequence 5, Application US/08440845D
; Patent No. 5955329
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; APPLICANT: Kridl, J.
; APPLICANT: Dehesh, K.
; APPLICANT: Knauf, V.
; TITLE OF INVENTION: Engineering Plant Thioesterases For
; TITLE OF INVENTION: Altered Substrate Specificity.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,845D
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-845D-5

Alignment Scores: 69 Length: 352

Pred. No.: 69 Matches: 8

Score: 8.00 Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.61% Indels: 0

DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-440-845D-5 (1-352)

QY 1382 CTCATCTCCTAAGTTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 4

US-08-868-458-5
; Sequence 5, Application US/08868458
; Patent No. 6150512
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; TITLE OF INVENTION: Engineering Plant Thioesterases And

US-08-868-458-5

Alignment Scores: 69 Length: 352

Pred. No.: 69 Matches: 8

Score: 8.00 Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.61% Indels: 0

DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-868-458-5 (1-352)

QY 1382 CTCATCTCCTAAGTTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 5

US-09-134-262-2
; Sequence 2, Application US/09134262
; Patent No. 6365802
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content in Soybean Oil
; FILE REFERENCE: Docket #17030
; CURRENT APPLICATION NUMBER: US/09/134,262
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: IBM PC; Windows NT 4.0; Microsoft Word For Windows 7.0a
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-134-262-2

Alignment Scores:

US-09-818-990B-1 (1-3963) x US-08-868-458-5 (1-352)

QY 1382 CTCATCTCCTAAGTTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 5

US-09-134-262-2
; Sequence 2, Application US/09134262
; Patent No. 6365802
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content in Soybean Oil
; FILE REFERENCE: Docket #17030
; CURRENT APPLICATION NUMBER: US/09/134,262
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: IBM PC; Windows NT 4.0; Microsoft Word For Windows 7.0a
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-134-262-2

Alignment Scores:

```

Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-134-262-2 (1-352)
QY 1382 CTCATCTCCTAGGTGAGTGGT 1405
Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 6
US-09-359-070-2
; Sequence 2, Application US/09359070
; Patent No. 6380462
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content
; FILE REFERENCE: 17030/01/US
; CURRENT APPLICATION NUMBER: US/09/359,070
; CURRENT FILING DATE: 1999-07-22
; PRIOR APPLICATION NUMBER: 09/134,262
; PRIOR FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-359-070-2

Alignment Scores:
Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-359-070-2 (1-352)
QY 1382 CTCATCTCCTAGGTGAGTGGT 1405
Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 7
US-08-464-523B-32
; Sequence 32, Application US/08464523B
; Patent No. 5723761
; GENERAL INFORMATION:
; APPLICANT: Toni A. Voelker
; APPLICANT: Ling Yuan
; APPLICANT: Jean Kridl
; APPLICANT: Deborah Hawkins
; APPLICANT: Aubrey Jones
; TITLE OF INVENTION: Plant Acyl ACP Thioesterase
; TITLE OF INVENTION: Sequences
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1(a)
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,523B
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13131
; FILING DATE: 10-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/152,004
; FILING DATE: 10-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,695
; FILING DATE: 16-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 100-1WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-464-523B-32

Alignment Scores:
Pred. No.: 68.8 Length: 362
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-464-523B-32 (1-362)
QY 1382 CTCATCTCCTAGGTGAGTGGT 1405
Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 8
US-08-948-176-25
; Sequence 25, Application US/08948176
; Patent No. 5945585
; GENERAL INFORMATION:
; APPLICANT: HITZ, WILLIAM D.
; APPLICANT: YADAV, NARENDRA S.
; TITLE OF INVENTION: ACYL-ACP THIOESTERASES GENES
; TITLE OF INVENTION: AND THEIR USE IN ALTERING PLANT
; TITLE OF INVENTION: OIL COMPOSITION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,176
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/631,264
```

;; FILING DATE: DECEMBER 20, 1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: CHRISTENBURY, LYNNE M.
;; REGISTRATION NUMBER: 30,971
;; REFERENCE/DOCKET NUMBER: CR-8926-C
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 302-992-5481
;; TELEFAX: 302-773-0164
;; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 362 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-948-176-25

Alignment Scores: Length: 362
Pred. No.: 68.8 Matches: 8
Score: 8.00 Conservative: 0
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.61%
Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-948-176-25 (1-362)

QY 1382 CTCATCTCTAAGTTGAGTGCT 1405
|||||
Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 9
US-08-440-845D-6
; Sequence 6, Application US/08440845D
; Patent No. 5955329
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; APPLICANT: Kridl, J.
; APPLICANT: Dehesh, K.
; APPLICANT: Knauf, V.
; TITLE OF INVENTION: Engineering Plant Thioesterases For
; TITLE OF INVENTION: Altered Substrate Specificity.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION NUMBER: US/08/440,845D
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 113
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-845D-6

Alignment Scores: Length: 362
Pred. No.: 68.8 Matches: 8
Score: 8.00 Conservative: 0
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.61%
Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-440-845D-6 (1-362)

QY 1382 CTCATCTCTAAGTTGAGTGCT 1405
|||||
Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 10
US-08-868-458-6
; Sequence 6, Application US/08868458
; Patent No. 6150512
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; TITLE OF INVENTION: Engineering Plant Thioesterases And
; TITLE OF INVENTION: Disclosure of Plant Thioesterases
; TITLE OF INVENTION: Having No. 6150512el Substrate Specificity
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
; COMPUTER: IBM PC
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: Microsoft Word For Window 95 7.0a
; CURRENT APPLICATION NUMBER: US/08/868,458
; FILING DATE: 03-Jun-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07064
; FILING DATE: 15-MAY-96
; APPLICATION NUMBER: 08/537,083
; FILING DATE: 29-SEPT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 133
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-868-458-6

Alignment Scores: Length: 362
Pred. No.: 68.8 Matches: 8
Score: 8.00 Conservative: 0
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.61%
Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-868-458-6 (1-362)

QY 1382 CTCCTCTCTTAAGTTGAGTGGT 1405
|||||
Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 11

US-09-303-592-2

; Sequence 2, Application US/09303592

; Patent No. 6426448

; GENERAL INFORMATION:

; APPLICANT: BOOTH, JOHN R.

; APPLICANT: BROGLIE, RICHARD M.

; APPLICANT: HITZ, WILLIAM D.

; APPLICANT: KINNEY, ANTHONY J.

; APPLICANT: KNOWLTON, SUSAN

; APPLICANT: SEBASTIAN, SCOTT A.

; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

; TITLE OF INVENTION: NOVEL GENE COMBINATIONS THAT ALTER THE QUALITY AND
; TITLE OF INVENTION: FUNCTIONALITY OF SOYBEAN OIL

; FILE REFERENCE: BB-1156

; CURRENT APPLICATION NUMBER: US/09/303,592

; EARLIER FILING DATE: 1999-05-03

; EARLIER APPLICATION NUMBER: 60/085,030

; EARLIER FILING DATE: MAY 11, 1998

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: MICROSOFT OFFICE 97

; SEQ ID NO 2

; LENGTH: 362

; TYPE: PRT

; ORGANISM: Brassica napus

US-09-303-592-2

Alignment Scores:

Pred. No.:	68.8	Length:	362
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.61%	Indels:	0
DB:	4	Gaps:	0

US-09-818-990B-1 (1-3963) x US-09-303-592-2 (1-362)

QY 1382 CTCCTCTCTTAAGTTGAGTGGT 1405

|||||

Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 12

US-09-303-592-4

; Sequence 4, Application US/09303592

; Patent No. 6426448

; GENERAL INFORMATION:

; APPLICANT: BOOTH, JOHN R.

; APPLICANT: BROGLIE, RICHARD M.

; APPLICANT: HITZ, WILLIAM D.

; APPLICANT: KINNEY, ANTHONY J.

; APPLICANT: KNOWLTON, SUSAN

; APPLICANT: SEBASTIAN, SCOTT A.

; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

; TITLE OF INVENTION: NOVEL GENE COMBINATIONS THAT ALTER THE QUALITY AND
; TITLE OF INVENTION: FUNCTIONALITY OF SOYBEAN OIL

; FILE REFERENCE: BB-1156

; CURRENT APPLICATION NUMBER: US/09/303,592

; EARLIER FILING DATE: 1999-05-03

; EARLIER APPLICATION NUMBER: 60/085,030

; EARLIER FILING DATE: MAY 11, 1998

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: MICROSOFT OFFICE 97

; SEQ ID NO 4

; LENGTH: 362

; TYPE: PRT

; ORGANISM: Brassica napus

US-09-303-592-4

Alignment Scores:

Pred. No.:	68.8	Length:	362
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.61%	Indels:	0
DB:	4	Gaps:	0

US-09-818-990B-1 (1-3963) x US-09-303-592-4 (1-362)

QY 1382 CTCCTCTCTTAAGTTGAGTGGT 1405

|||||

Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 13

US-09-108-020-54

; Sequence 54, Application US/09108020A

; Patent No. 6143561

; GENERAL INFORMATION:

; APPLICANT: Randall, Douglas D.

; APPLICANT: Johnston, Mark L.

; APPLICANT: Mierzyk, Jan A.

; APPLICANT: Mooney, Brian P.

; APPLICANT: Luehly, Michael H.

; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND

; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO

; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANOATE BIOSYNTHESIS IN PLANTS

; FILE REFERENCE: UMO 1482

; CURRENT APPLICATION NUMBER: US/09/108,020A

; CURRENT FILING DATE: 1998-06-30

; EARLIER APPLICATION NUMBER: 60/051,291

; EARLIER FILING DATE: 1997-06-30

; EARLIER APPLICATION NUMBER: 60/055,255

; EARLIER FILING DATE: 1997-08-01

; EARLIER APPLICATION NUMBER: 60/076,544

; EARLIER FILING DATE: 1998-03-02

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 54

; LENGTH: 375

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: consensus

US-09-108-020-54

Alignment Scores:

Pred. No.:	68.4	Length:	375
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.63%	Indels:	0
DB:	4	Gaps:	0

US-09-818-990B-1 (1-3963) x US-09-108-020-54 (1-375)

QY 2289 GAGGCTTCTTTCCTCACTGCTCT 2266

|||||

Db 188 GluAlaPhePheAlaHisCysPro 195

RESULT 14

US-09-459-133-13

; Sequence 13, Application US/09459133

; Patent No. 6416988

; GENERAL INFORMATION:

; APPLICANT: Conklin, Darrell C.

; APPLICANT: Yamamoto, Gayle

; APPLICANT: Jaspers, Stephen R.

; APPLICANT: Gao, Zeren

; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS

; FILE REFERENCE: 98-77

; CURRENT APPLICATION NUMBER: US/09/459,133

; CURRENT FILING DATE: 1999-12-10

; PRIOR APPLICATION NUMBER: 60/111,697

; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-459-133-13

Alignment Scores:
Pred. No.: 68.1 Length: 389
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.63%
Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-459-133-13 (1-389)

QY 1568 GTCGCTATTGTGCTGCAGTA 1545
|||||
Db 139 ValProTyrLeuLeuAlaVal 146

RESULT 15

US-09-108-020-53
; Sequence 53, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luethy, Michael H.
; APPLICANT: Mooney, Brian P.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANOATE BIOSYNTHESIS IN PLANTS
; FILE REFERENCE: UMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Bovine
US-09-108-020-53

Alignment Scores:
Pred. No.: 68 Length: 391
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.63%
Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-108-020-53 (1-391)

QY 2289 GAGCGTTCTTGTCTACTGTCCT 2266
|||||
Db 200 GluAlaPheAlaHisCysPro 207

Search completed: November 30, 2002, 18:51:23
Job time : 49 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:48:35 ; Search time 16.5 Seconds

(without alignments)

7649.408 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacagacagcataga.....tggagagtgatgaactttaa 3963

Scoring table:

OLIGO

Xgapop 60.0 , Xgapext 60.0

Ygapop 60.0 , Ygapext 60.0

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 102317 seqs, 15924203 residues

Word size: 1

Total number of hits satisfying chosen parameters: 183048

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh

-Q/cgn2_1/USFTO.spool/US09818990/runat_26112002_093224_22184/app_query.fasta_1.4103

-DB=PublishedApplications_AA -QMT=fastan -SUFFIX=oligo.rapb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo

-ALIGN=human40.cdi -LIST=1000 -DOALIGN=200 -THR SCORE=quality -THR_MIN=1

-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09818990.cgn_1.1.10_arunat_26112002_093224_22184

-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*

7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261	19.8	264	10	US-09-764-853-638
2	113	8.6	159	10	US-09-764-853-817
3	19	1.4	544	10	US-09-925-299-1004
4	8	0.6	17	10	US-09-864-761-45408
					Sequence 638, App
					Sequence 817, App
					Sequence 1004, Ap
					Sequence 45408, A

46	10	US-09-864-761-38762	0.6	8	Sequence 38762, A
51	10	US-09-864-761-39357	0.6	8	Sequence 39357, A
73	10	US-09-867-550-1662	0.6	8	Sequence 1662, Ap
94	10	US-09-858-664A-23	0.6	8	Sequence 23, Appl
119	10	US-09-858-664A-30	0.6	8	Sequence 30, Appl
257	10	US-09-925-302-757	0.6	8	Sequence 757, App
339	10	US-09-925-297-706	0.6	8	Sequence 706, App
352	10	US-09-134-262-2	0.6	8	Sequence 2, Appli
378	10	US-09-801-368-330	0.6	8	Sequence 330, App
595	10	US-09-815-242-12107	0.6	8	Sequence 12107, A
646	10	US-09-777-710A-1	0.6	8	Sequence 1, Appli
670	10	US-09-777-710A-15	0.6	8	Sequence 15, Appl
746	9	US-10-106-092-2	0.6	8	Sequence 2, Appli
848	10	US-09-839-185-8	0.6	8	Sequence 8, Appli
1286	9	US-10-017-216-7	0.6	8	Sequence 7, Appli
1597	9	US-10-017-216-6	0.6	8	Sequence 6, Appli
1641	9	US-10-017-216-5	0.6	8	Sequence 5, Appli
2053	9	US-10-017-216-2	0.6	8	Sequence 2, Appli
2054	12	US-10-028-946-2	0.6	8	Sequence 2, Appli
2055	9	US-10-017-216-4	0.6	8	Sequence 4, Appli
3594	10	US-09-911-842-4	0.6	8	Sequence 4, Appli
10	10	US-09-884-441-427	0.5	7	Sequence 427, App
11	8	US-08-424-550B-521	0.5	7	Sequence 521, App
21	10	US-09-884-441-402	0.5	7	Sequence 402, App
24	10	US-09-864-761-44088	0.5	7	Sequence 44088, A
25	10	US-09-864-761-44777	0.5	7	Sequence 44777, A
28	10	US-09-982-172-226	0.5	7	Sequence 226, App
30	10	US-09-864-761-47695	0.5	7	Sequence 47695, A
34	12	US-10-001-870-183	0.5	7	Sequence 183, App
37	10	US-09-982-172-94	0.5	7	Sequence 94, Appl
39	10	US-09-864-761-42302	0.5	7	Sequence 42302, A
41	9	US-09-989-919-120	0.5	7	Sequence 120, App
42	10	US-09-726-643-147	0.5	7	Sequence 147, App
42	10	US-09-864-761-45179	0.5	7	Sequence 45179, A
47	10	US-09-864-761-44811	0.5	7	Sequence 44811, A
48	10	US-09-864-761-41726	0.5	7	Sequence 41726, A
51	10	US-09-864-761-47816	0.5	7	Sequence 47816, A
52	10	US-09-864-761-35437	0.5	7	Sequence 35437, A
53	10	US-09-864-761-35128	0.5	7	Sequence 35128, A
61	10	US-09-864-761-41115	0.5	7	Sequence 41115, A
68	10	US-09-925-300-1560	0.5	7	Sequence 1560, Ap
72	10	US-09-891-171-11	0.5	7	Sequence 11, Appl
74	10	US-09-764-853-448	0.5	7	Sequence 448, App
78	9	US-09-749-637A-399	0.5	7	Sequence 399, App
78	10	US-09-925-297-748	0.5	7	Sequence 748, App
89	10	US-09-864-761-43218	0.5	7	Sequence 43218, A
93	10	US-09-864-761-38395	0.5	7	Sequence 38395, A
95	10	US-09-841-879B-12	0.5	7	Sequence 12, Appl
98	9	US-10-103-511-8	0.5	7	Sequence 8, Appli
98	10	US-09-867-550-442	0.5	7	Sequence 442, App
98	10	US-09-731-872-343	0.5	7	Sequence 343, App
98	10	US-09-805-204-8	0.5	7	Sequence 8, Appli
101	9	US-09-286-488-68	0.5	7	Sequence 68, Appl
101	10	US-09-737-178-68	0.5	7	Sequence 68, Appl
105	9	US-09-286-488-69	0.5	7	Sequence 69, Appl
105	10	US-09-737-178-69	0.5	7	Sequence 69, Appl
110	10	US-09-765-272-102	0.5	7	Sequence 102, App
110	10	US-09-800-723-108	0.5	7	Sequence 108, App
112	10	US-09-891-171-7	0.5	7	Sequence 7, Appli
116	9	US-09-896-738-7	0.5	7	Sequence 7, Appli
116	10	US-09-850-165-82	0.5	7	Sequence 82, Appl
116	10	US-09-850-165-111	0.5	7	Sequence 111, App
119	10	US-09-867-550-246	0.5	7	Sequence 246, App
119	10	US-09-731-872-319	0.5	7	Sequence 319, App
120	10	US-09-764-877-1102	0.5	7	Sequence 1102, Ap
120	12	US-10-096-703-3	0.5	7	Sequence 3, Appli
121	9	US-09-852-797-68	0.5	7	Sequence 68, Appl
121	9	US-09-852-797-85	0.5	7	Sequence 85, Appl
121	10	US-09-853-161-68	0.5	7	Sequence 68, Appl
121	10	US-09-853-161-85	0.5	7	Sequence 85, Appl
121	10	US-09-852-659A-68	0.5	7	Sequence 68, Appl
121	10	US-09-852-659A-85	0.5	7	Sequence 85, Appl
121	12	US-10-115-406-18	0.5	7	Sequence 18, Appl

c 78	Sequence 44, Appl	c 151	7	0.5	279	10	US-09-741-152-4	Sequence 4, Appl
c 79	Sequence 22, Appl	c 152	7	0.5	279	10	US-09-874-585B-34	Sequence 34, Appl
c 80	Sequence 18, Appl	c 153	7	0.5	280	10	US-09-925-300-1411	Sequence 1411, Ap
c 81	Sequence 997, App	c 154	7	0.5	280	12	US-10-001-843-145	Sequence 145, App
c 82	Sequence 201, App	c 155	7	0.5	282	9	US-09-992-598-291	Sequence 21, App
c 83	Sequence 31, Appl	c 156	7	0.5	282	9	US-09-896-738-2	Sequence 2, Appl
c 84	Sequence 31, Appl	c 157	7	0.5	282	9	US-09-915-789A-5	Sequence 5, Appl
c 85	Sequence 359, App	c 158	7	0.5	282	10	US-09-778-320-208	Sequence 208, App
c 86	Sequence 359, App	c 159	7	0.5	282	10	US-09-850-178-33	Sequence 33, Appl
c 87	Sequence 359, App	c 160	7	0.5	282	10	US-09-877-065-8	Sequence 8, Appl
c 88	Sequence 359, App	c 161	7	0.5	282	10	US-09-989-722-291	Sequence 291, App
c 89	Sequence 359, App	c 162	7	0.5	282	10	US-09-989-723-291	Sequence 291, App
c 90	Sequence 359, App	c 163	7	0.5	282	10	US-09-989-727-291	Sequence 291, App
c 91	Sequence 359, App	c 164	7	0.5	282	10	US-09-989-727-291	Sequence 291, App
c 92	Sequence 359, App	c 165	7	0.5	282	10	US-09-910-689-208	Sequence 208, App
c 93	Sequence 359, App	c 166	7	0.5	282	10	US-09-989-731-291	Sequence 291, App
c 94	Sequence 359, App	c 167	7	0.5	282	10	US-09-884-441-333	Sequence 333, App
c 95	Sequence 359, App	c 168	7	0.5	282	10	US-09-989-732-291	Sequence 291, App
c 96	Sequence 359, App	c 169	7	0.5	282	10	US-09-991-073-291	Sequence 291, App
c 97	Sequence 359, App	c 170	7	0.5	282	10	US-09-990-442-291	Sequence 291, App
c 98	Sequence 444, App	c 171	7	0.5	282	10	US-09-991-163-291	Sequence 291, App
c 99	Sequence 431, App	c 172	7	0.5	282	10	US-09-993-604-291	Sequence 291, App
c 100	Sequence 4, Appl	c 173	7	0.5	282	10	US-09-990-456-291	Sequence 291, App
c 101	Sequence 26, Appl	c 174	7	0.5	282	10	US-09-989-721-291	Sequence 291, App
c 102	Sequence 164, App	c 175	7	0.5	282	12	US-10-006-867-60	Sequence 60, Appl
c 103	Sequence 77, Appl	c 176	7	0.5	282	12	US-10-052-586-218	Sequence 218, App
c 104	Sequence 1154, Ap	c 177	7	0.5	282	12	US-10-010-742-208	Sequence 208, App
c 105	Sequence 78, Appl	c 178	7	0.5	306	9	US-09-712-363-186	Sequence 186, App
c 106	Sequence 183, App	c 179	7	0.5	307	9	US-09-764-868-1100	Sequence 1100, Ap
c 107	Sequence 34202, A	c 180	7	0.5	309	10	US-09-778-320-209	Sequence 209, App
c 108	Sequence 34988, A	c 181	7	0.5	309	10	US-09-910-689-209	Sequence 209, App
c 109	Sequence 36844, A	c 182	7	0.5	309	10	US-09-884-441-332	Sequence 392, App
c 110	Sequence 731, App	c 183	7	0.5	309	12	US-10-010-742-209	Sequence 209, App
c 111	Sequence 39, Appl	c 184	7	0.5	310	10	US-09-286-488-19	Sequence 19, Appl
c 112	Sequence 1104, Ap	c 185	7	0.5	310	10	US-09-737-178-19	Sequence 19, Appl
c 113	Sequence 101, App	c 186	7	0.5	310	10	US-09-925-300-1602	Sequence 1602, Ap
c 114	Sequence 2, Appl	c 187	7	0.5	312	10	US-09-925-302-783	Sequence 783, App
c 115	Sequence 19, Appl	c 188	7	0.5	314	10	US-09-768-826-43	Sequence 43, Appl
c 116	Sequence 10, Appl	c 189	7	0.5	315	10	US-09-815-242-11636	Sequence 11636, A
c 117	Sequence 20, Appl	c 190	7	0.5	319	10	US-09-925-302-758	Sequence 758, App
c 118	Sequence 10825, A	c 191	7	0.5	319	10	US-09-886-055-95	Sequence 95, Appl
c 119	Sequence 677, App	c 192	7	0.5	325	10	US-09-927-113-10	Sequence 10, Appl
c 120	Sequence 18, Appl	c 193	7	0.5	331	9	US-09-941-831-17	Sequence 17, Appl
c 121	Sequence 11, Appl	c 194	7	0.5	338	10	US-09-741-669-365	Sequence 365, App
c 122	Sequence 817, App	c 195	7	0.5	338	10	US-09-912-020-373	Sequence 373, App
c 123	Sequence 5404, Ap	c 196	7	0.5	340	10	US-09-815-242-13257	Sequence 13257, A
c 124	Sequence 12578, A	c 197	7	0.5	340	10	US-09-815-242-13579	Sequence 13579, A
c 125	Sequence 12778, A	c 198	7	0.5	349	10	US-09-919-497-76	Sequence 76, Appl
c 126	Sequence 54, Appl	c 199	7	0.5	352	9	US-09-104-063-4	Sequence 4, Appl
c 127	Sequence 68, Appl	c 200	7	0.5	352	10	US-09-953-692-2	Sequence 2, Appl
c 128	Sequence 74, Appl	c 201	7	0.5	352	10	US-09-953-717-2	Sequence 2, Appl
c 129	Sequence 488, App	c 202	7	0.5	354	10	US-09-925-302-501	Sequence 501, App
c 130	Sequence 1616, App	c 203	7	0.5	356	10	US-09-925-300-1326	Sequence 1326, Ap
c 131	Sequence 19, Appl	c 204	7	0.5	367	9	US-09-286-488-20	Sequence 20, Appl
c 132	Sequence 1678, Ap	c 205	7	0.5	367	9	US-09-813-358-18	Sequence 18, Appl
c 133	Sequence 95, Appl	c 206	7	0.5	367	10	US-09-737-178-20	Sequence 20, Appl
c 134	Sequence 10415, A	c 207	7	0.5	369	9	US-09-895-913A-180	Sequence 180, App
c 135	Sequence 3, Appl	c 208	7	0.5	371	10	US-09-764-853-600	Sequence 600, App
c 136	Sequence 17, Appl	c 209	7	0.5	371	10	US-09-764-853-790	Sequence 790, App
c 137	Sequence 27, Appl	c 210	7	0.5	372	10	US-09-768-894A-2	Sequence 2, Appl
c 138	Sequence 18, Appl	c 211	7	0.5	380	10	US-09-815-242-11249	Sequence 11249, A
c 139	Sequence 143, App	c 212	7	0.5	414	9	US-10-108-605-87	Sequence 1027, Ap
c 140	Sequence 18, Appl	c 213	7	0.5	415	9	US-09-764-868-1027	Sequence 87, Appl
c 141	Sequence 446, App	c 214	7	0.5	427	9	US-09-975-253-2	Sequence 2, Appl
c 142	Sequence 3, Appl	c 215	7	0.5	433	10	US-09-801-368-38	Sequence 38, Appl
c 143	Sequence 12, Appl	c 216	7	0.5	434	10	US-09-801-574-78	Sequence 78, Appl
c 144	Sequence 2, Appl	c 217	7	0.5	448	10	US-09-815-242-12017	Sequence 12017, A
c 145	Sequence 2, Appl	c 218	7	0.5	450	10	US-09-765-111A-14	Sequence 14, Appl
c 146	Sequence 2, Appl	c 219	7	0.5	453	10	US-09-815-242-10872	Sequence 10872, A
c 147	Sequence 5, Appl	c 220	7	0.5	458	10	US-09-826-463-1	GENERAL INFORMA
c 148	Sequence 14, Appl	c 221	7	0.5	459	10	US-09-771-161A-198	Sequence 198, App
c 149	Sequence 640, App	c 222	7	0.5	474	9	US-09-998-909-14	Sequence 14, Appl
c 150	Sequence 34, Appl	c 223	7	0.5	476	10	US-09-925-300-1582	Sequence 1582, Ap

c 224	7	0.5	487	10	US-09-825-414-66	Sequence 66, Appl	297	7	0.5	1193	12	US-10-071-900-4	Sequence 4, Appl
c 225	7	0.5	491	10	US-09-815-242-10409	Sequence 10409, A	c 298	7	0.5	1206	12	US-10-032-717-2	Sequence 2, Appl
c 226	7	0.5	491	10	US-09-815-242-14033	Sequence 14033, A	c 299	7	0.5	1207	12	US-09-927-112-2	Sequence 2, Appl
c 227	7	0.5	500	10	US-09-848-726-2	Sequence 2, Appl	c 300	7	0.5	1210	12	US-10-032-717-4	Sequence 4, Appl
c 228	7	0.5	504	10	US-09-801-368-208	Sequence 208, App	301	7	0.5	1214	9	US-09-945-901-54	Sequence 54, Appl
c 229	7	0.5	507	10	US-09-729-674-14	Sequence 14, Appl	302	7	0.5	1214	9	US-10-007-747-54	Sequence 54, Appl
c 230	7	0.5	511	10	US-09-912-020-259	Sequence 259, App	303	7	0.5	1219	9	US-09-945-901-50	Sequence 50, Appl
c 231	7	0.5	522	10	US-09-765-272-120	Sequence 120, App	304	7	0.5	1219	9	US-10-007-747-50	Sequence 50, Appl
c 232	7	0.5	529	10	US-09-954-314-22	Sequence 22, Appl	305	7	0.5	1231	9	US-09-945-901-48	Sequence 48, Appl
c 233	7	0.5	530	9	US-10-108-605-171	Sequence 171, Appl	306	7	0.5	1231	9	US-10-007-747-48	Sequence 48, Appl
c 234	7	0.5	541	10	US-09-981-421-4	Sequence 4, Appl	307	7	0.5	1236	9	US-09-945-901-6	Sequence 6, Appl
c 235	7	0.5	541	12	US-10-157-447-2	Sequence 2, Appl	308	7	0.5	1236	9	US-10-007-747-6	Sequence 6, Appl
c 236	7	0.5	542	10	US-09-925-301-1004	Sequence 1004, Ap	309	7	0.5	1239	9	US-09-945-901-52	Sequence 52, Appl
c 237	7	0.5	551	10	US-09-815-242-10500	Sequence 10500, A	310	7	0.5	1239	9	US-10-007-747-52	Sequence 52, Appl
c 238	7	0.5	554	10	US-09-815-242-12033	Sequence 12033, A	311	7	0.5	1244	9	US-09-945-901-46	Sequence 46, Appl
c 239	7	0.5	555	10	US-09-912-020-308	Sequence 308, App	312	7	0.5	1244	9	US-10-007-747-46	Sequence 46, Appl
c 240	7	0.5	555	10	US-09-815-242-10477	Sequence 10477, A	c 313	7	0.5	1260	9	US-10-072-621-7	Sequence 27, Appl
c 241	7	0.5	555	10	US-09-815-242-11735	Sequence 11735, A	c 314	7	0.5	1280	9	US-10-024-918-27	Sequence 27, Appl
c 242	7	0.5	555	10	US-09-815-242-13709	Sequence 13709, A	c 315	7	0.5	1280	10	US-09-866-866A-2	Sequence 2, Appl
c 243	7	0.5	576	10	US-09-854-731-19	Sequence 19, Appl	c 316	7	0.5	1280	10	US-09-866-866A-4	Sequence 4, Appl
c 244	7	0.5	582	10	US-09-919-497-100	Sequence 100, App	317	7	0.5	1360	10	US-09-871-916-2	Sequence 2, Appl
c 245	7	0.5	597	10	US-09-828-313-38	Sequence 38, Appl	318	7	0.5	1369	10	US-09-729-674-42	Sequence 42, Appl
c 246	7	0.5	599	10	US-09-815-242-11326	Sequence 11326, A	319	7	0.5	1463	9	US-09-971-536-69	Sequence 69, Appl
c 247	7	0.5	623	10	US-09-854-731-4	Sequence 4, Appl	320	7	0.5	1523	12	US-10-011-064-2	Sequence 2, Appl
c 248	7	0.5	625	10	US-09-854-731-18	Sequence 18, Appl	321	7	0.5	1523	12	US-10-052-586-230	Sequence 290, App
c 249	7	0.5	628	9	US-09-906-209-8	Sequence 8, Appl	322	7	0.5	1614	9	US-09-887-540A-2	Sequence 2, Appl
c 250	7	0.5	643	10	US-09-371-900-39	Sequence 39, Appl	c 323	7	0.5	1659	10	US-09-801-368-118	Sequence 118, App
c 251	7	0.5	643	10	US-09-924-417-56	Sequence 56, Appl	c 324	7	0.5	2012	9	US-09-808-602-68	Sequence 68, Appl
c 252	7	0.5	643	12	US-10-063-763-11	Sequence 11, Appl	c 325	7	0.5	2150	9	US-10-135-322-17	Sequence 17, Appl
c 253	7	0.5	655	10	US-09-815-242-11751	Sequence 11751, A	c 326	7	0.5	2783	10	US-09-816-669A-14	Sequence 14, Appl
c 254	7	0.5	658	10	US-09-815-242-10947	Sequence 10947, A	c 327	7	0.5	2783	10	US-09-816-669A-14	Sequence 14, Appl
c 255	7	0.5	663	9	US-10-108-605-245	Sequence 245, App	328	7	0.5	2923	10	US-09-788-711A-4	Sequence 4, Appl
c 256	7	0.5	666	10	US-09-765-272-2	Sequence 2, Appl	329	7	0.5	2956	10	US-09-788-711A-2	Sequence 2, Appl
c 257	7	0.5	670	12	US-10-063-763-10	Sequence 10, Appl	c 330	7	0.5	3313	10	US-09-737-149-29	Sequence 29, Appl
c 258	7	0.5	671	12	US-10-052-586-346	Sequence 346, App	331	7	0.5	7968	9	US-10-077-130-5	Sequence 5, Appl
c 259	7	0.5	679	12	US-10-119-714-1	Sequence 1, Appl	c 332	6	0.5	9	1	US-08-821-739A-72	Sequence 72, Appl
c 260	7	0.5	719	10	US-09-815-242-13423	Sequence 13423, A	c 333	6	0.5	9	10	US-09-753-831-7	Sequence 7, Appl
c 261	7	0.5	742	10	US-09-732-180-9	Sequence 9, Appl	c 334	6	0.5	11	12	US-10-007-761-70	Sequence 70, Appl
c 262	7	0.5	750	10	US-09-815-242-13406	Sequence 9, Appl	c 335	6	0.5	11	12	US-10-007-761-71	Sequence 71, Appl
c 263	7	0.5	760	10	US-09-764-367A-9	Sequence 9, Appl	c 336	6	0.5	12	10	US-09-873-459A-13	Sequence 13, Appl
c 264	7	0.5	764	10	US-09-828-466-4	Sequence 4, Appl	c 337	6	0.5	12	10	US-09-873-459A-14	Sequence 14, Appl
c 265	7	0.5	764	10	US-09-828-466-5	Sequence 5, Appl	c 338	6	0.5	12	10	US-09-873-459A-16	Sequence 16, Appl
c 266	7	0.5	769	12	US-10-087-110-6	Sequence 6, Appl	339	6	0.5	13	10	US-09-938-315-83	Sequence 83, Appl
c 267	7	0.5	770	10	US-09-801-368-32	Sequence 32, Appl	c 340	6	0.5	14	10	US-09-874-470-10	Sequence 10, Appl
c 268	7	0.5	797	10	US-09-795-232-2	Sequence 2, Appl	341	6	0.5	15	10	US-09-758-318-19	Sequence 19, Appl
c 269	7	0.5	805	10	US-09-799-777-19	Sequence 19, Appl	c 342	6	0.5	15	10	US-09-785-802A-4	Sequence 4, Appl
c 270	7	0.5	807	10	US-09-822-295-2	Sequence 2, Appl	c 343	6	0.5	17	10	US-09-839-447A-82	Sequence 82, Appl
c 271	7	0.5	810	10	US-09-815-242-5386	Sequence 5386, Ap	c 344	6	0.5	18	10	US-09-864-761-39272	Sequence 39272, A
c 272	7	0.5	818	10	US-09-815-242-12424	Sequence 12424, A	345	6	0.5	18	10	US-09-734-520-23	Sequence 23, Appl
c 273	7	0.5	818	10	US-09-815-242-12789	Sequence 12789, A	346	6	0.5	18	12	US-10-012-034A-23	Sequence 11, Appl
c 274	7	0.5	875	10	US-09-848-035-16	Sequence 16, Appl	c 347	6	0.5	19	9	US-09-291-925A-11	Sequence 44921, A
c 275	7	0.5	875	10	US-09-986-224-16	Sequence 16, Appl	348	6	0.5	19	10	US-09-864-761-44921	Sequence 44921, A
c 276	7	0.5	875	12	US-10-087-110-2	Sequence 2, Appl	349	6	0.5	19	10	US-09-864-761-48622	Sequence 48622, A
c 277	7	0.5	890	12	US-10-028-056-3	Sequence 3, Appl	c 350	6	0.5	19	10	US-09-976-787-32	Sequence 32, Appl
c 278	7	0.5	891	9	US-09-976-059-18	Sequence 18, Appl	c 351	6	0.5	19	10	US-09-976-787-34	Sequence 34, Appl
c 279	7	0.5	912	9	US-10-047-542-75	Sequence 75, Appl	c 352	6	0.5	19	10	US-09-865-198-33	Sequence 33, Appl
c 280	7	0.5	917	9	US-10-047-542-76	Sequence 76, Appl	c 353	6	0.5	19	10	US-09-987-456-146	Sequence 146, App
c 281	7	0.5	917	9	US-10-047-542-87	Sequence 87, Appl	c 354	6	0.5	19	12	US-10-153-064-71	Sequence 71, Appl
c 282	7	0.5	924	9	US-10-047-542-67	Sequence 67, Appl	c 355	6	0.5	20	10	US-09-864-761-36422	Sequence 36422, A
c 283	7	0.5	952	9	US-09-924-396B-18	Sequence 18, Appl	c 356	6	0.5	20	10	US-09-864-761-43265	Sequence 43265, A
c 284	7	0.5	978	10	US-09-864-761-43207	Sequence 43207, A	c 357	6	0.5	20	10	US-09-813-333-75	Sequence 75, Appl
c 285	7	0.5	978	10	US-09-893-817-4	Sequence 4, Appl	358	6	0.5	21	10	US-09-864-761-34628	Sequence 34628, A
c 286	7	0.5	978	10	US-09-893-817-8	Sequence 8, Appl	359	6	0.5	21	10	US-09-864-761-43818	Sequence 43818, A
c 287	7	0.5	985	9	US-09-978-295A-211	Sequence 211, App	c 360	6	0.5	22	9	US-09-880-132-41	Sequence 41, Appl
c 288	7	0.5	985	9	US-09-978-697-211	Sequence 211, App	361	6	0.5	22	10	US-09-815-156-13	Sequence 13, Appl
c 289	7	0.5	1004	10	US-09-767-215-2	Sequence 2, Appl	362	6	0.5	22	10	US-09-864-761-42735	Sequence 42735, A
c 290	7	0.5	1040	10	US-09-765-272-118	Sequence 118, Appl	363	6	0.5	22	10	US-09-864-761-46647	Sequence 46647, A
c 291	7	0.5	1055	10	US-09-759-667A-3	Sequence 3, Appl	364	6	0.5	22	10	US-09-835-232-10	Sequence 10, Appl
c 292	7	0.5	1065	10	US-09-771-161A-239	Sequence 239, App	c 365	6	0.5	22	10	US-09-965-099-39	Sequence 39, Appl
c 293	7	0.5	1138	10	US-09-767-215-5	Sequence 5, Appl	c 366	6	0.5	22	10	US-09-880-149-41	Sequence 41, Appl
c 294	7	0.5	1148	10	US-09-815-242-11895	Sequence 11895, A	c 367	6	0.5	22	12	US-10-051-852-39	Sequence 39, Appl
c 295	7	0.5	1156	10	US-09-757-716-1	Sequence 1, Appl	c 368	6	0.5	23	9	US-09-989-920-269	Sequence 269, App
c 296	7	0.5	1163	10	US-09-350-259-4	Sequence 4, Appl	c 369	6	0.5	23	10	US-09-785-632A-147	Sequence 147, App

c 370	6	0.5	23	10	US-09-965-099-24	Sequence 24, Appl	443	37	10	US-09-764-869-661	Sequence 661, App
c 371	6	0.5	23	10	US-09-965-099-76	Sequence 76, Appl	c 444	38	9	US-09-915-676-3	Sequence 3, Appl1
c 372	6	0.5	23	12	US-10-051-852-24	Sequence 24, Appl	c 445	38	10	US-09-466-320-2	Sequence 2, Appl1
c 373	6	0.5	23	12	US-10-051-852-76	Sequence 76, Appl	446	38	10	US-09-864-761-33587	Sequence 33587, A
c 374	6	0.5	24	10	US-09-775-925-4	Sequence 4, Appl1	447	38	10	US-09-864-761-33287	Sequence 36287, A
c 375	6	0.5	24	10	US-09-864-761-33869	Sequence 33869, A	448	38	10	US-09-864-761-41336	Sequence 41336, A
c 376	6	0.5	24	10	US-09-864-761-40814	Sequence 40814, A	c 449	38	10	US-09-864-761-41879	Sequence 41879, A
c 377	6	0.5	24	10	US-09-864-761-46886	Sequence 46886, A	450	38	10	US-09-864-761-42714	Sequence 42714, A
c 378	6	0.5	24	10	US-09-864-761-47163	Sequence 47163, A	451	38	10	US-09-864-761-45152	Sequence 45152, A
c 379	6	0.5	24	10	US-09-847-519A-4	Sequence 4, Appl1	452	38	10	US-09-864-761-45542	Sequence 45542, A
c 380	6	0.5	24	10	US-09-964-277-17	Sequence 17, Appl	453	38	10	US-09-864-761-46762	Sequence 46762, A
c 381	6	0.5	25	10	US-09-205-658-228	Sequence 228, App	c 454	38	10	US-09-764-847-940	Sequence 940, App
c 382	6	0.5	25	10	US-09-864-761-36459	Sequence 36459, A	c 455	39	10	US-09-864-761-33428	Sequence 33428, A
c 383	6	0.5	25	10	US-09-864-761-44796	Sequence 44796, A	456	39	10	US-09-864-761-42104	Sequence 42104, A
c 384	6	0.5	25	10	US-09-864-761-48129	Sequence 48129, A	457	39	10	US-09-864-761-45869	Sequence 45869, A
c 385	6	0.5	25	12	US-10-066-895-17	Sequence 17, Appl	c 458	39	10	US-09-925-297-756	Sequence 756, App
c 386	6	0.5	26	10	US-09-864-761-39591	Sequence 39591, A	459	40	10	US-09-843-845-7	Sequence 7, Appl1
c 387	6	0.5	26	10	US-09-864-761-45669	Sequence 45669, A	c 460	40	10	US-09-864-761-34778	Sequence 34778, A
c 388	6	0.5	26	10	US-09-864-761-47507	Sequence 47507, A	c 461	40	10	US-09-864-761-45519	Sequence 45519, A
c 389	6	0.5	26	12	US-10-066-895-15	Sequence 15, Appl	c 462	40	10	US-09-864-761-45963	Sequence 45963, A
c 390	6	0.5	27	9	US-09-880-132-14	Sequence 14, Appl	c 463	40	10	US-09-925-299-1452	Sequence 1452, Ap
c 391	6	0.5	27	10	US-09-880-149-14	Sequence 14, Appl	464	40	10	US-09-764-869-774	Sequence 774, App
c 392	6	0.5	28	10	US-09-864-761-40492	Sequence 40492, A	465	40	10	US-09-764-846-227	Sequence 227, App
c 393	6	0.5	28	10	US-09-864-761-44946	Sequence 44946, A	466	41	10	US-09-864-761-40140	Sequence 40140, A
c 394	6	0.5	28	10	US-09-864-761-48752	Sequence 48752, A	c 467	41	10	US-09-814-122-46	Sequence 46, Appl
c 395	6	0.5	28	10	US-09-071-838-97	Sequence 97, Appl	c 468	41	10	US-09-815-242-10672	Sequence 10672, A
c 396	6	0.5	29	10	US-09-799-983-21	Sequence 21, Appl	c 469	41	10	US-09-939-980-290	Sequence 290, App
c 397	6	0.5	29	10	US-09-864-761-37579	Sequence 37579, A	c 470	42	10	US-09-864-761-36082	Sequence 36082, A
c 398	6	0.5	29	10	US-09-864-761-41049	Sequence 41049, A	c 471	42	10	US-09-864-761-42202	Sequence 42202, A
c 399	6	0.5	29	10	US-09-864-761-42321	Sequence 42321, A	472	42	10	US-09-864-761-46395	Sequence 46395, A
c 400	6	0.5	30	10	US-09-864-761-35699	Sequence 35699, A	473	42	10	US-09-864-761-46732	Sequence 46732, A
c 401	6	0.5	30	10	US-09-929-315-14	Sequence 14, Appl	474	43	9	US-09-981-876-185	Sequence 185, App
c 402	6	0.5	30	10	US-09-929-315-17	Sequence 17, Appl	475	43	9	US-09-981-876-205	Sequence 205, App
c 403	6	0.5	30	10	US-09-950-773-8	Sequence 8, Appl1	476	43	10	US-09-864-761-38836	Sequence 38836, A
c 404	6	0.5	30	10	US-09-993-215-14	Sequence 14, Appl	c 477	43	10	US-09-864-761-48891	Sequence 48891, A
c 405	6	0.5	30	10	US-09-993-215-17	Sequence 17, Appl	478	43	10	US-09-864-761-34453	Sequence 34453, A
c 406	6	0.5	31	10	US-09-810-310-7	Sequence 7, Appl1	c 479	44	10	US-09-864-761-42506	Sequence 42506, A
c 407	6	0.5	31	10	US-09-810-310-14	Sequence 14, Appl	480	44	10	US-09-764-877-1452	Sequence 1452, Ap
c 408	6	0.5	31	10	US-09-864-761-39495	Sequence 39495, A	c 481	45	10	US-09-205-658-62	Sequence 62, Appl
c 409	6	0.5	31	10	US-09-864-761-41712	Sequence 41712, A	c 482	45	10	US-09-205-658-92	Sequence 92, Appl
c 410	6	0.5	31	10	US-09-764-869-1011	Sequence 1011, Ap	c 483	45	10	US-09-205-658-205	Sequence 205, Appl
c 411	6	0.5	32	10	US-09-864-761-36013	Sequence 36013, A	c 484	45	10	US-09-844-353A-62	Sequence 62, Appl
c 412	6	0.5	32	10	US-09-864-761-37002	Sequence 37002, A	c 485	45	10	US-09-844-353A-92	Sequence 92, Appl
c 413	6	0.5	32	10	US-09-864-761-44308	Sequence 44308, A	c 486	45	10	US-09-864-761-36504	Sequence 36504, A
c 414	6	0.5	32	10	US-09-864-761-47615	Sequence 47615, A	487	45	10	US-09-864-761-43881	Sequence 43881, A
c 415	6	0.5	32	10	US-09-764-847-650	Sequence 650, App	488	45	10	US-09-864-761-45540	Sequence 45540, A
c 416	6	0.5	33	10	US-09-864-761-42161	Sequence 42161, A	c 489	45	10	US-09-864-761-47862	Sequence 47862, A
c 417	6	0.5	33	10	US-09-864-761-46877	Sequence 46877, A	490	46	8	US-08-424-5508-286	Sequence 256, App
c 418	6	0.5	33	12	US-10-001-870-162	Sequence 162, App	c 491	46	10	US-09-764-887-283	Sequence 283, App
c 419	6	0.5	34	10	US-09-864-761-38045	Sequence 38045, A	492	46	10	US-09-864-761-36968	Sequence 36968, A
c 420	6	0.5	34	10	US-09-864-761-43787	Sequence 43787, A	493	46	10	US-09-864-761-37165	Sequence 37165, A
c 421	6	0.5	34	10	US-09-864-761-44326	Sequence 44326, A	c 494	46	10	US-09-864-761-38126	Sequence 38126, A
c 422	6	0.5	34	10	US-09-864-761-45067	Sequence 45067, A	c 495	46	10	US-09-864-761-40103	Sequence 40103, A
c 423	6	0.5	34	10	US-09-864-761-45367	Sequence 45367, A	c 496	46	10	US-09-864-761-46281	Sequence 46281, A
c 424	6	0.5	34	10	US-09-281-717-25	Sequence 25, Appl	c 497	46	10	US-09-764-869-806	Sequence 806, App
c 425	6	0.5	35	10	US-09-864-761-36426	Sequence 36426, A	c 498	47	10	US-09-739-254-109	Sequence 109, App
c 426	6	0.5	35	10	US-09-864-761-40258	Sequence 40258, A	c 499	47	10	US-09-904-615-109	Sequence 109, App
c 427	6	0.5	35	10	US-09-864-761-41118	Sequence 41118, A	c 500	47	10	US-09-864-761-38801	Sequence 38801, A
c 428	6	0.5	36	8	US-08-424-5508-414	Sequence 414, App	501	47	10	US-09-864-761-42877	Sequence 42877, A
c 429	6	0.5	36	9	US-10-420-319-19	Sequence 19, Appl	502	47	10	US-09-864-761-45534	Sequence 45534, A
c 430	6	0.5	36	10	US-09-758-318-20	Sequence 20, Appl	503	47	10	US-09-864-761-47530	Sequence 47530, A
c 431	6	0.5	36	10	US-09-864-761-37869	Sequence 37869, A	504	47	10	US-09-864-761-47530	Sequence 1434, Ap
c 432	6	0.5	36	10	US-09-864-761-41511	Sequence 41511, A	505	47	10	US-09-925-299-1434	Sequence 507, App
c 433	6	0.5	36	10	US-09-733-524-12	Sequence 12, Appl	506	47	10	US-09-764-877-1213	Sequence 1213, Ap
c 434	6	0.5	37	9	US-09-915-676-2	Sequence 2, Appl1	507	48	10	US-09-864-761-34146	Sequence 34146, A
c 435	6	0.5	37	10	US-09-466-320-1	Sequence 1, Appl1	508	48	10	US-09-864-761-36278	Sequence 36278, A
c 436	6	0.5	37	10	US-09-726-643-174	Sequence 174, App	c 509	48	10	US-09-864-761-37629	Sequence 37629, A
c 437	6	0.5	37	10	US-09-864-761-39522	Sequence 39522, A	c 510	48	10	US-09-864-761-39535	Sequence 39535, A
c 438	6	0.5	37	10	US-09-864-761-40909	Sequence 40909, A	c 511	48	10	US-09-864-761-43879	Sequence 43879, A
c 439	6	0.5	37	10	US-09-864-761-41082	Sequence 41082, A	c 512	48	10	US-09-864-761-45635	Sequence 45635, A
c 440	6	0.5	37	10	US-09-864-761-41269	Sequence 41269, A	513	48	10	US-09-864-761-47978	Sequence 47978, A
c 441	6	0.5	37	10	US-09-864-761-45320	Sequence 45320, A	514	49	9	US-09-836-392-34	Sequence 34, Appl
c 442	6	0.5	37	10	US-09-864-761-47992	Sequence 47992, A	515	49	10	US-09-739-907-73	Sequence 73, Appl

c 516	6	0.5	49	10	US-09-739-254-116	Sequence 116, App	589	6	0.5	60	10	US-09-864-761-46151	Sequence 46151, A
c 517	6	0.5	49	10	US-09-904-615-116	Sequence 116, App	590	6	0.5	60	10	US-09-789-561-107	Sequence 107, App
c 518	6	0.5	49	10	US-09-864-761-33750	Sequence 33750, A	591	6	0.5	60	10	US-09-789-561-145	Sequence 145, App
c 519	6	0.5	49	10	US-09-864-761-36944	Sequence 36944, A	c 592	6	0.5	60	10	US-09-867-550-520	Sequence 520, App
c 520	6	0.5	49	10	US-09-864-761-40297	Sequence 40297, A	c 593	6	0.5	60	12	US-10-001-843-191	Sequence 191, App
c 521	6	0.5	49	10	US-09-864-761-46765	Sequence 46765, A	594	6	0.5	61	10	US-09-864-761-34440	Sequence 34440, A
c 522	6	0.5	49	12	US-10-001-843-180	Sequence 180, App	c 595	6	0.5	61	10	US-09-864-761-43093	Sequence 43093, A
c 523	6	0.5	50	10	US-09-864-761-35577	Sequence 35577, A	c 596	6	0.5	61	10	US-09-867-550-1192	Sequence 1192, App
c 524	6	0.5	50	10	US-09-864-761-38636	Sequence 38636, A	597	6	0.5	61	10	US-09-867-550-1802	Sequence 1802, App
c 525	6	0.5	50	10	US-09-864-761-44405	Sequence 44405, A	598	6	0.5	62	10	US-09-864-761-35144	Sequence 35144, A
c 526	6	0.5	50	10	US-09-764-869-985	Sequence 985, App	c 599	6	0.5	62	10	US-09-864-761-39904	Sequence 39904, A
c 527	6	0.5	50	12	US-10-139-876-9	Sequence 9, Appl	600	6	0.5	62	10	US-09-864-761-40967	Sequence 40967, A
c 528	6	0.5	51	9	US-09-989-919-85	Sequence 85, Appl	c 601	6	0.5	63	10	US-09-741-669-317	Sequence 317, App
c 529	6	0.5	51	10	US-09-864-761-42096	Sequence 42096, A	c 602	6	0.5	63	10	US-09-912-020-384	Sequence 384, App
c 530	6	0.5	51	10	US-09-864-761-46820	Sequence 46820, A	c 603	6	0.5	63	10	US-09-864-761-35276	Sequence 35276, A
c 531	6	0.5	51	10	US-09-764-869-1080	Sequence 1080, App	c 604	6	0.5	63	10	US-09-864-761-37584	Sequence 37584, A
c 532	6	0.5	51	10	US-09-867-550-1332	Sequence 1332, App	c 605	6	0.5	63	10	US-09-815-242-10351	Sequence 10351, A
c 533	6	0.5	51	10	US-09-764-847-510	Sequence 510, App	c 606	6	0.5	63	10	US-09-867-550-184	Sequence 184, App
c 534	6	0.5	51	10	US-09-925-300-1252	Sequence 1252, App	c 607	6	0.5	63	10	US-09-764-877-1602	Sequence 1602, App
c 535	6	0.5	52	10	US-09-864-761-37266	Sequence 37266, A	c 608	6	0.5	64	8	US-08-424-5508-384	Sequence 384, App
c 536	6	0.5	52	10	US-09-864-761-40193	Sequence 40193, A	609	6	0.5	64	9	US-09-886-156-39	Sequence 39, Appl
c 537	6	0.5	52	10	US-09-864-761-41707	Sequence 41707, A	610	6	0.5	64	9	US-09-886-150-39	Sequence 39, Appl
c 538	6	0.5	53	10	US-09-864-761-39665	Sequence 39665, A	611	6	0.5	64	10	US-09-864-761-43911	Sequence 43911, A
c 539	6	0.5	53	10	US-09-864-761-41739	Sequence 41739, A	612	6	0.5	64	10	US-09-764-869-1099	Sequence 1099, App
c 540	6	0.5	53	10	US-09-864-761-42900	Sequence 42900, A	c 613	6	0.5	64	10	US-09-745-763-52	Sequence 52, Appl
c 541	6	0.5	53	10	US-09-864-761-48757	Sequence 48757, A	614	6	0.5	64	10	US-09-915-582-86	Sequence 86, Appl
c 542	6	0.5	54	10	US-09-764-887-268	Sequence 268, App	c 615	6	0.5	65	10	US-09-466-320-12	Sequence 12, Appl
c 543	6	0.5	54	10	US-09-864-761-41834	Sequence 41834, A	c 616	6	0.5	65	10	US-09-864-761-40730	Sequence 40730, A
c 544	6	0.5	54	10	US-09-864-761-43863	Sequence 43863, A	c 617	6	0.5	66	10	US-09-466-320-13	Sequence 13, Appl
c 545	6	0.5	54	10	US-09-864-761-46759	Sequence 46759, A	c 618	6	0.5	66	10	US-09-864-761-35267	Sequence 35267, A
c 546	6	0.5	54	10	US-09-796-858-6	Sequence 6, Appl	c 619	6	0.5	66	10	US-09-864-761-35847	Sequence 35847, A
c 547	6	0.5	54	10	US-09-764-847-862	Sequence 862, App	c 620	6	0.5	66	10	US-09-864-761-36613	Sequence 36613, A
c 548	6	0.5	54	10	US-09-905-176-6	Sequence 6, Appl	c 621	6	0.5	66	10	US-09-864-761-36864	Sequence 36864, A
c 549	6	0.5	54	10	US-09-905-176-7	Sequence 7, Appl	c 622	6	0.5	66	10	US-09-864-761-38038	Sequence 38038, A
c 550	6	0.5	54	10	US-09-905-176-8	Sequence 8, Appl	c 623	6	0.5	66	10	US-09-864-761-42501	Sequence 42501, A
c 551	6	0.5	55	9	US-09-970-966-200	Sequence 200, App	c 624	6	0.5	66	10	US-09-864-761-43212	Sequence 43212, A
c 552	6	0.5	55	10	US-09-825-294-200	Sequence 200, App	625	6	0.5	66	10	US-09-864-761-45142	Sequence 45142, A
c 553	6	0.5	55	10	US-09-759-143-911	Sequence 911, App	c 626	6	0.5	66	10	US-09-864-761-48141	Sequence 48141, A
c 554	6	0.5	55	10	US-09-864-761-38937	Sequence 38937, A	c 627	6	0.5	66	10	US-09-864-761-48359	Sequence 48359, A
c 555	6	0.5	55	10	US-09-864-761-48486	Sequence 48486, A	c 628	6	0.5	66	12	US-10-001-843-126	Sequence 126, App
c 556	6	0.5	55	10	US-09-780-669-911	Sequence 911, App	629	6	0.5	67	10	US-09-764-877-1827	Sequence 1827, App
c 557	6	0.5	55	10	US-09-822-817-911	Sequence 911, App	c 630	6	0.5	68	10	US-09-220-920-50	Sequence 50, Appl
c 558	6	0.5	55	10	US-09-867-550-1962	Sequence 1962, App	c 631	6	0.5	68	10	US-09-466-320-11	Sequence 11, Appl
c 559	6	0.5	56	9	US-10-001-835-142	Sequence 142, App	c 632	6	0.5	68	10	US-09-864-761-36782	Sequence 36782, A
c 560	6	0.5	56	10	US-09-864-761-34621	Sequence 34621, A	c 633	6	0.5	68	10	US-09-864-761-38672	Sequence 38672, A
c 561	6	0.5	56	10	US-09-864-761-38631	Sequence 38631, A	c 634	6	0.5	68	10	US-09-864-761-40131	Sequence 40131, A
c 562	6	0.5	56	10	US-09-864-761-42647	Sequence 42647, A	c 635	6	0.5	68	10	US-09-864-761-40644	Sequence 40644, A
c 563	6	0.5	56	10	US-09-799-514-17	Sequence 17, Appl	636	6	0.5	68	10	US-09-864-761-46215	Sequence 46215, A
c 564	6	0.5	56	10	US-09-764-877-1122	Sequence 1122, App	637	6	0.5	68	10	US-09-864-761-47335	Sequence 47335, A
c 565	6	0.5	57	10	US-09-864-761-40354	Sequence 40354, A	638	6	0.5	68	10	US-09-864-761-48597	Sequence 48597, A
c 566	6	0.5	57	10	US-09-864-761-42360	Sequence 42360, A	639	6	0.5	68	10	US-09-815-242-13205	Sequence 13205, A
c 567	6	0.5	57	10	US-09-864-761-44565	Sequence 44565, A	640	6	0.5	68	10	US-09-867-550-876	Sequence 876, App
c 568	6	0.5	57	10	US-09-864-761-46064	Sequence 46064, A	c 641	6	0.5	68	10	US-09-867-550-876	Sequence 876, App
c 569	6	0.5	57	10	US-09-864-761-46311	Sequence 46311, A	642	6	0.5	9	US-09-970-033-10	Sequence 10, Appl	
c 570	6	0.5	57	10	US-09-864-761-47667	Sequence 47667, A	c 643	6	0.5	69	10	US-09-864-761-36291	Sequence 36291, A
c 571	6	0.5	57	10	US-09-864-761-48589	Sequence 48589, App	c 644	6	0.5	69	10	US-09-925-299-946	Sequence 946, App
c 572	6	0.5	58	10	US-09-739-254-98	Sequence 98, Appl	c 645	6	0.5	69	10	US-09-764-877-1649	Sequence 1649, App
c 573	6	0.5	58	10	US-09-904-615-98	Sequence 98, Appl	646	6	0.5	70	9	US-09-970-033-6	Sequence 6, Appl
c 574	6	0.5	58	10	US-09-864-761-36415	Sequence 36415, A	c 647	6	0.5	70	9	US-10-010-408-9	Sequence 9, Appl
c 575	6	0.5	58	10	US-09-864-761-37226	Sequence 37226, A	648	6	0.5	70	10	US-09-864-761-35373	Sequence 35373, A
c 576	6	0.5	58	10	US-09-864-761-37378	Sequence 37378, A	649	6	0.5	70	10	US-09-864-761-41240	Sequence 41240, A
c 577	6	0.5	58	10	US-09-864-761-46792	Sequence 46792, A	c 650	6	0.5	70	10	US-09-764-860-366	Sequence 366, App
c 578	6	0.5	58	10	US-09-922-261-173	Sequence 173, App	c 651	6	0.5	70	10	US-09-764-847-722	Sequence 722, App
c 579	6	0.5	59	10	US-09-726-643-124	Sequence 124, App	c 652	6	0.5	71	10	US-09-864-761-43147	Sequence 43147, A
c 580	6	0.5	59	10	US-09-864-761-39622	Sequence 39622, A	c 653	6	0.5	71	10	US-09-815-242-11044	Sequence 11044, A
c 581	6	0.5	59	10	US-09-864-761-47424	Sequence 47424, A	654	6	0.5	72	10	US-09-827-949-17	Sequence 17, Appl
c 582	6	0.5	59	10	US-09-925-301-1141	Sequence 1141, App	c 655	6	0.5	72	10	US-09-764-887-231	Sequence 231, App
c 583	6	0.5	59	10	US-09-764-877-1231	Sequence 1231, App	c 656	6	0.5	72	10	US-09-764-877-1398	Sequence 1398, App
c 584	6	0.5	59	12	US-10-001-843-181	Sequence 181, App	c 657	6	0.5	73	8	US-08-424-5508-510	Sequence 510, App
c 585	6	0.5	60	10	US-09-864-761-37221	Sequence 37221, A	c 658	6	0.5	73	10	US-09-864-761-40477	Sequence 40477, A
c 586	6	0.5	60	10	US-09-864-761-42942	Sequence 42942, A	c 659	6	0.5	73	10	US-09-864-761-44311	Sequence 44311, A
c 587	6	0.5	60	10	US-09-864-761-43672	Sequence 43672, A	c 660	6	0.5	73	10	US-09-864-761-44846	Sequence 44846, A
c 588	6	0.5	60	10	US-09-864-761-43672	Sequence 43672, A	661	6	0.5	73	10	US-09-864-761-47336	Sequence 47336, A

c 562	73	10	US-09-864-761-47461	Sequence 47461, A	735	6	0.5	88	10	US-09-864-761-35206	Sequence 35206, A
c 563	73	10	US-09-764-846-242	Sequence 242, App	736	6	0.5	88	10	US-09-864-761-36391	Sequence 36391, A
c 564	74	10	US-09-864-761-43233	Sequence 43233, A	737	6	0.5	88	10	US-09-864-761-37179	Sequence 37179, A
c 565	74	10	US-09-864-761-47476	Sequence 47476, A	c 738	6	0.5	88	10	US-09-925-299-827	Sequence 827, App
c 566	74	10	US-09-864-761-47793	Sequence 47793, A	739	6	0.5	88	10	US-09-867-550-740	Sequence 740, App
c 567	74	10	US-09-864-761-48313	Sequence 48313, A	740	6	0.5	88	10	US-09-751-100B-64	Sequence 64, Appl
c 568	74	10	US-09-764-877-1892	Sequence 1892, App	c 741	6	0.5	89	9	US-10-138-787-13	Sequence 13, Appl
c 569	75	10	US-09-864-761-33418	Sequence 33418, A	c 742	6	0.5	89	10	US-09-864-761-46146	Sequence 46146, A
c 570	75	10	US-09-864-761-36328	Sequence 36328, A	c 743	6	0.5	89	10	US-09-815-242-5867	Sequence 5867, Ap
c 571	75	10	US-09-864-761-44757	Sequence 44757, A	c 744	6	0.5	89	10	US-09-764-855-99	Sequence 99, Appl
c 572	75	10	US-09-864-761-48587	Sequence 48587, A	c 745	6	0.5	89	10	US-09-862-179A-17	Sequence 17, Appl
c 573	76	9	US-09-957-995A-12	Sequence 12, Appl	c 746	6	0.5	90	10	US-09-287-070-11	Sequence 11, Appl
c 574	76	10	US-09-908-711-77	Sequence 77, Appl	c 747	6	0.5	90	10	US-09-864-761-33620	Sequence 33620, A
c 575	76	10	US-09-864-761-34053	Sequence 34053, A	c 748	6	0.5	90	10	US-09-864-761-36923	Sequence 36923, A
c 576	76	10	US-09-864-761-39315	Sequence 39315, A	c 749	6	0.5	90	10	US-09-864-761-43004	Sequence 43004, A
c 577	77	9	US-10-001-835-181	Sequence 181, App	c 750	6	0.5	91	9	US-09-843-676-220	Sequence 220, App
c 578	77	10	US-09-864-761-40922	Sequence 40922, A	c 751	6	0.5	91	9	US-09-897-878B-9	Sequence 9, Appl1
c 579	77	10	US-09-864-761-41086	Sequence 41086, A	c 752	6	0.5	91	9	US-09-984-245-152	Sequence 152, App
c 580	77	10	US-09-864-761-46163	Sequence 46163, A	c 753	6	0.5	91	10	US-09-864-761-38421	Sequence 38421, A
c 581	77	10	US-09-764-860-372	Sequence 372, App	c 754	6	0.5	91	10	US-09-864-761-43733	Sequence 43733, A
c 582	78	10	US-09-864-761-35930	Sequence 35930, A	c 755	6	0.5	92	9	US-09-897-878B-6	Sequence 6, Appl1
c 583	78	10	US-09-864-761-47381	Sequence 47381, A	c 756	6	0.5	92	9	US-10-054-873-2	Sequence 2, Appl1
c 584	78	10	US-09-867-550-1970	Sequence 1970, App	c 757	6	0.5	92	10	US-09-993-844-21	Sequence 21, Appl
c 585	78	10	US-09-731-872-368	Sequence 368, App	c 758	6	0.5	92	10	US-09-354-453-49	Sequence 49, Appl
c 586	78	10	US-09-764-877-1056	Sequence 1056, App	c 759	6	0.5	93	10	US-09-287-849-35	Sequence 35, Appl
c 587	79	10	US-09-864-761-48376	Sequence 48376, A	c 760	6	0.5	93	10	US-09-908-711-99	Sequence 99, Appl
c 588	79	10	US-09-764-869-681	Sequence 681, App	c 761	6	0.5	93	10	US-09-864-761-39557	Sequence 39557, A
c 589	79	12	US-10-001-843-148	Sequence 148, App	c 762	6	0.5	93	10	US-09-864-761-46583	Sequence 46583, A
c 590	80	10	US-09-864-761-48470	Sequence 48470, A	c 763	6	0.5	93	10	US-09-867-550-488	Sequence 448, App
c 591	80	10	US-09-867-550-1924	Sequence 1924, App	c 764	6	0.5	93	10	US-09-867-550-590	Sequence 590, App
c 592	81	10	US-09-764-870-515	Sequence 515, App	c 765	6	0.5	93	10	US-09-731-872-465	Sequence 1977, Ap
c 593	81	10	US-09-864-761-34936	Sequence 34936, A	c 766	6	0.5	93	12	US-09-764-877-1977	Sequence 22, Appl
c 594	81	10	US-09-864-761-41593	Sequence 41593, A	c 767	6	0.5	94	9	US-10-014-269-2	Sequence 215, App
c 595	81	10	US-09-864-761-46205	Sequence 46205, A	c 768	6	0.5	94	10	US-09-843-676-215	Sequence 1343, Ap
c 596	81	10	US-09-864-761-48698	Sequence 48698, A	c 769	6	0.5	94	10	US-09-925-301-1343	Sequence 31, Appl
c 597	81	10	US-09-730-617-91	Sequence 91, Appl	c 770	6	0.5	94	10	US-09-728-721-31	Sequence 5, Appl1
c 598	81	10	US-09-731-872-281	Sequence 281, App	c 771	6	0.5	94	10	US-09-886-426-5	Sequence 31, Appl1
c 599	81	10	US-09-764-877-2046	Sequence 2046, App	c 772	6	0.5	94	10	US-09-739-707-2	Sequence 2, Appl1
c 700	81	12	US-10-052-586-432	Sequence 432, App	c 773	6	0.5	94	12	US-10-105-931-31	Sequence 31, Appl1
c 701	82	10	US-09-864-761-33885	Sequence 33885, A	c 774	6	0.5	95	10	US-09-925-302-450	Sequence 450, App
c 702	82	10	US-09-864-761-41506	Sequence 41506, A	c 775	6	0.5	95	10	US-09-864-761-47792	Sequence 47792, A
c 703	82	10	US-09-864-761-46565	Sequence 46565, A	c 776	6	0.5	95	10	US-09-815-242-10524	Sequence 10524, A
c 704	82	10	US-09-925-301-1467	Sequence 1467, App	c 777	6	0.5	95	10	US-09-867-550-720	Sequence 720, App
c 705	82	10	US-09-815-242-14090	Sequence 14090, A	c 778	6	0.5	95	10	US-09-731-872-429	Sequence 429, App
c 706	82	10	US-09-867-550-492	Sequence 492, App	c 779	6	0.5	95	10	US-09-925-300-1824	Sequence 1824, Ap
c 707	82	10	US-09-764-877-1582	Sequence 1582, App	c 780	6	0.5	95	10	US-09-071-838-273	Sequence 273, App
c 708	83	10	US-09-864-761-33875	Sequence 33875, A	c 781	6	0.5	96	10	US-09-864-761-39732	Sequence 39732, A
c 709	83	10	US-09-864-761-48130	Sequence 48130, A	c 782	6	0.5	96	10	US-09-864-761-42326	Sequence 42326, A
c 710	83	10	US-09-864-761-48233	Sequence 48233, A	c 783	6	0.5	96	10	US-09-754-997A-20	Sequence 20, Appl
c 711	83	10	US-09-939-980-410	Sequence 410, App	c 784	6	0.5	96	10	US-09-263-959-1201	Sequence 1201, Ap
c 712	83	10	US-09-764-864-1432	Sequence 1432, App	c 785	6	0.5	96	10	US-09-925-300-1090	Sequence 1090, Ap
c 713	84	9	US-10-001-835-160	Sequence 160, App	c 786	6	0.5	96	12	US-10-023-528-26	Sequence 26, Appl
c 714	84	10	US-09-864-761-38458	Sequence 38458, A	c 787	6	0.5	97	10	US-09-925-302-804	Sequence 804, App
c 715	85	9	US-10-067-989-10	Sequence 10, Appl	c 788	6	0.5	97	10	US-09-864-761-43929	Sequence 43929, A
c 716	85	10	US-09-764-868-1203	Sequence 1203, App	c 789	6	0.5	97	10	US-09-887-586A-40	Sequence 40, Appl
c 717	85	10	US-09-815-242-4934	Sequence 4934, App	c 790	6	0.5	97	10	US-09-903-012-40	Sequence 12, Appl
c 718	85	10	US-09-764-860-559	Sequence 559, App	c 791	6	0.5	98	9	US-09-843-676-12	Sequence 48859, A
c 719	85	10	US-09-764-847-839	Sequence 839, App	c 792	6	0.5	98	10	US-09-864-761-48869	Sequence 1765, Ap
c 720	86	10	US-09-864-761-40312	Sequence 40312, A	c 793	6	0.5	99	10	US-09-764-877-1765	Sequence 35148, A
c 721	86	10	US-09-764-869-935	Sequence 935, App	c 794	6	0.5	99	10	US-09-864-761-35148	Sequence 43778, A
c 722	86	10	US-09-764-846-196	Sequence 196, App	c 795	6	0.5	99	10	US-09-864-761-43778	Sequence 46653, A
c 723	86	10	US-09-995-494-106	Sequence 106, App	c 796	6	0.5	99	10	US-09-864-761-46653	Sequence 1480, Ap
c 724	86	10	US-09-764-877-1630	Sequence 1630, App	c 797	6	0.5	99	10	US-09-925-299-1480	Sequence 172, App
c 725	87	9	US-09-852-797-89	Sequence 89, Appl	c 798	6	0.5	100	9	US-10-001-835-172	Sequence 820, App
c 726	87	9	US-09-989-920-203	Sequence 203, App	c 799	6	0.5	100	9	US-09-764-868-820	Sequence 191, App
c 727	87	10	US-09-220-920-115	Sequence 115, App	c 800	6	0.5	100	9	US-09-989-920-191	Sequence 264, App
c 728	87	10	US-09-864-761-33727	Sequence 33727, A	c 801	6	0.5	100	9	US-09-989-920-264	Sequence 67, Appl
c 729	87	10	US-09-864-761-34744	Sequence 34744, A	c 802	6	0.5	100	10	US-09-739-907-67	Sequence 35700, A
c 730	87	10	US-09-864-761-41037	Sequence 41037, A	c 803	6	0.5	100	10	US-09-864-761-35700	Sequence 36185, A
c 731	87	10	US-09-853-161-89	Sequence 89, Appl	c 804	6	0.5	100	10	US-09-864-761-36185	Sequence 22, App
c 732	87	10	US-09-852-659A-89	Sequence 89, Appl	c 805	6	0.5	100	10	US-09-764-898-222	Sequence 17, Appl
c 733	87	10	US-09-867-550-1468	Sequence 1468, App	c 806	6	0.5	100	10	US-09-885-478-17	
c 734	88	10	US-09-764-887-294	Sequence 294, App	c 807	6	0.5	100	10		

808	6	0.5	100	10	US-09-922-261-62	Sequence 62, Appl	881	6	0.5	113	12	US-10-029-654-4	Sequence 4, Appli
809	6	0.5	100	10	US-09-872-523-8	Sequence 8, Appli	c 882	6	0.5	114	10	US-09-864-761-35473	Sequence 35473, A
c 810	6	0.5	100	12	US-10-052-586-526	Sequence 526, App	c 883	6	0.5	114	10	US-09-263-959-300	Sequence 300, App
c 811	6	0.5	100	12	US-10-040-916-42	Sequence 42, Appl	c 884	6	0.5	115	10	US-09-860-232A-13	Sequence 13, Appl
c 812	6	0.5	101	8	US-08-424-550B-60	Sequence 60, Appli	c 885	6	0.5	115	10	US-09-925-302-597	Sequence 597, App
c 813	6	0.5	102	9	US-10-004-381-8	Sequence 8, Appli	c 886	6	0.5	115	10	US-09-864-761-34042	Sequence 34042, A
c 814	6	0.5	102	10	US-09-864-761-34648	Sequence 34648, A	c 887	6	0.5	115	10	US-09-922-261-58	Sequence 58, Appl
c 815	6	0.5	102	10	US-09-864-761-37445	Sequence 37445, A	c 888	6	0.5	115	10	US-09-922-261-58	Sequence 59, Appl
c 816	6	0.5	102	10	US-09-864-761-42384	Sequence 42384, A	c 889	6	0.5	115	10	US-09-995-494-79	Sequence 79, Appl
c 817	6	0.5	103	10	US-09-925-299-1497	Sequence 1497, Ap	c 890	6	0.5	116	9	US-09-263-959-315	Sequence 315, App
c 818	6	0.5	103	10	US-09-867-550-1044	Sequence 1044, Ap	c 891	6	0.5	116	10	US-10-001-835-161	Sequence 161, App
c 819	6	0.5	103	10	US-09-942-447-8	Sequence 8, Appli	c 892	6	0.5	116	10	US-09-864-761-40290	Sequence 40290, A
c 820	6	0.5	103	10	US-09-751-100B-76	Sequence 76, Appl	c 893	6	0.5	116	10	US-09-789-561-178	Sequence 178, App
c 821	6	0.5	104	10	US-09-764-887-260	Sequence 260, App	c 894	6	0.5	116	10	US-09-867-550-146	Sequence 146, App
c 822	6	0.5	104	10	US-09-764-870-461	Sequence 461, App	c 895	6	0.5	116	10	US-09-764-853-739	Sequence 739, App
c 823	6	0.5	104	10	US-09-864-761-35266	Sequence 35266, A	c 896	6	0.5	117	8	US-09-967-552A-42	Sequence 42, Appl
c 824	6	0.5	104	10	US-09-864-761-39527	Sequence 39527, A	c 897	6	0.5	117	8	US-08-790-540A-2	Sequence 2, Appli
c 825	6	0.5	104	10	US-09-864-761-41522	Sequence 41522, A	c 898	6	0.5	117	8	US-08-790-540A-6	Sequence 6, Appli
c 826	6	0.5	104	10	US-09-864-761-41522	Sequence 41522, A	c 899	6	0.5	117	8	US-08-791-391A-2	Sequence 2, Appli
c 827	6	0.5	104	10	US-09-815-242-5168	Sequence 5168, Ap	c 899	6	0.5	117	8	US-08-791-391A-6	Sequence 6, Appli
c 828	6	0.5	104	10	US-09-740-668A-46	Sequence 46, Appl	c 900	6	0.5	117	9	US-09-852-209A-19	Sequence 19, Appl
c 829	6	0.5	104	12	US-10-052-586-550	Sequence 550, App	c 901	6	0.5	117	9	US-09-950-933A-62	Sequence 62, Appl
c 830	6	0.5	105	9	US-09-897-878B-8	Sequence 8, Appli	c 902	6	0.5	117	9	US-10-091-236-8	Sequence 8, Appli
c 831	6	0.5	105	10	US-09-864-761-38933	Sequence 38933, A	c 903	6	0.5	117	10	US-09-925-300-958	Sequence 958, App
c 832	6	0.5	105	10	US-09-864-761-45587	Sequence 45587, A	c 904	6	0.5	118	10	US-09-799-777-31	Sequence 31, Appl
c 833	6	0.5	105	10	US-09-939-980-467	Sequence 467, App	c 905	6	0.5	118	10	US-09-771-161A-97	Sequence 97, Appl
c 834	6	0.5	105	10	US-09-764-853-744	Sequence 744, App	c 906	6	0.5	119	10	US-09-764-870-429	Sequence 429, App
c 835	6	0.5	105	10	US-09-764-853-86	Sequence 86, Appl	c 907	6	0.5	119	10	US-09-912-020-279	Sequence 279, App
c 836	6	0.5	105	10	US-09-764-864-1226	Sequence 1226, Ap	c 908	6	0.5	119	10	US-09-867-550-1676	Sequence 1676, Ap
c 837	6	0.5	105	10	US-09-751-100B-84	Sequence 84, Appl	c 909	6	0.5	119	10	US-09-850-163-54	Sequence 54, Appl
c 838	6	0.5	106	9	US-09-897-878B-5	Sequence 5, Appli	c 910	6	0.5	119	10	US-09-925-300-997	Sequence 997, App
c 839	6	0.5	106	9	US-09-950-933A-70	Sequence 70, Appl	c 911	6	0.5	120	9	US-09-860-670-97	Sequence 97, Appl
c 840	6	0.5	106	10	US-09-925-302-640	Sequence 640, App	c 912	6	0.5	120	10	US-09-764-870-480	Sequence 480, App
c 841	6	0.5	106	10	US-09-864-761-36805	Sequence 36805, A	c 913	6	0.5	120	10	US-09-864-761-35797	Sequence 35797, A
c 842	6	0.5	106	10	US-09-764-869-1058	Sequence 1058, Ap	c 914	6	0.5	120	10	US-09-767-041-51	Sequence 51, Appl
c 843	6	0.5	106	10	US-09-815-242-11082	Sequence 11082, A	c 915	6	0.5	120	10	US-09-867-550-28	Sequence 28, Appl
c 844	6	0.5	106	10	US-09-925-297-734	Sequence 734, App	c 916	6	0.5	121	10	US-09-764-864-1211	Sequence 1211, Ap
c 845	6	0.5	106	10	US-09-764-877-1980	Sequence 1980, Ap	c 917	6	0.5	121	10	US-09-864-761-43304	Sequence 43304, A
c 846	6	0.5	106	10	US-09-925-300-1749	Sequence 1749, Ap	c 918	6	0.5	121	10	US-09-925-301-1246	Sequence 1246, Ap
c 847	6	0.5	106	10	US-09-071-838-208	Sequence 208, App	c 919	6	0.5	121	10	US-09-925-301-1309	Sequence 1309, Ap
c 848	6	0.5	107	9	US-09-764-904-62	Sequence 62, Appl	c 920	6	0.5	121	10	US-09-764-860-505	Sequence 505, App
c 849	6	0.5	107	10	US-09-220-920-52	Sequence 52, Appl	c 921	6	0.5	121	10	US-09-893-737-182	Sequence 182, App
c 850	6	0.5	107	10	US-09-925-301-1608	Sequence 1608, Ap	c 922	6	0.5	122	10	US-09-815-242-11697	Sequence 11697, A
c 851	6	0.5	107	10	US-09-764-860-510	Sequence 510, App	c 923	6	0.5	122	10	US-09-771-161A-162	Sequence 162, App
c 852	6	0.5	107	10	US-09-764-860-580	Sequence 580, App	c 924	6	0.5	122	10	US-09-771-161A-162	Sequence 162, App
c 853	6	0.5	107	12	US-09-738-973-1113	Sequence 113, App	c 925	6	0.5	123	9	US-09-950-933A-71	Sequence 71, Appl
c 854	6	0.5	108	9	US-10-025-687-14	Sequence 14, Appl	c 926	6	0.5	123	9	US-09-981-876-187	Sequence 187, App
c 855	6	0.5	108	9	US-09-970-033-9	Sequence 9, Appli	c 927	6	0.5	123	9	US-09-981-876-243	Sequence 243, App
c 856	6	0.5	108	9	US-10-001-835-201	Sequence 201, App	c 928	6	0.5	123	10	US-09-216-393-120	Sequence 120, App
c 857	6	0.5	108	9	US-09-950-933A-61	Sequence 61, Appl	c 929	6	0.5	123	10	US-09-820-893-81	Sequence 81, Appl
c 858	6	0.5	108	10	US-09-822-263-14	Sequence 14, Appl	c 930	6	0.5	123	10	US-09-820-893-92	Sequence 92, Appl
c 859	6	0.5	108	10	US-09-822-263-14	Sequence 14, Appl	c 931	6	0.5	123	10	US-09-881-752A-324	Sequence 324, App
c 860	6	0.5	108	10	US-09-771-161A-179	Sequence 179, App	c 932	6	0.5	123	10	US-09-925-300-1696	Sequence 2, Appli
c 861	6	0.5	109	10	US-09-764-877-1708	Sequence 1708, Ap	c 933	6	0.5	124	9	US-09-897-878B-2	Sequence 266, App
c 862	6	0.5	109	10	US-09-751-100B-96	Sequence 96, Appl	c 934	6	0.5	124	10	US-09-731-872-266	Sequence 316, App
c 863	6	0.5	110	9	US-09-764-877-1287	Sequence 1287, Ap	c 935	6	0.5	124	12	US-09-893-737-316	Sequence 10, Appl
c 864	6	0.5	110	9	US-09-950-933A-69	Sequence 69, Appl	c 936	6	0.5	124	12	US-10-066-895-10	Sequence 454, App
c 865	6	0.5	110	10	US-09-815-242-5029	Sequence 5029, Ap	c 937	6	0.5	125	9	US-09-978-295A-454	Sequence 454, App
c 866	6	0.5	110	10	US-09-789-561-90	Sequence 90, Appl	c 938	6	0.5	125	9	US-09-978-697-454	Sequence 103, App
c 867	6	0.5	110	10	US-09-764-878-125	Sequence 125, App	c 939	6	0.5	125	9	US-09-852-797-103	Sequence 36, Appl
c 868	6	0.5	110	10	US-09-764-864-1057	Sequence 1057, Ap	c 940	6	0.5	125	10	US-09-866-562-36	Sequence 1155, Ap
c 869	6	0.5	111	10	US-09-729-674-277	Sequence 277, App	c 941	6	0.5	125	10	US-09-925-301-1155	Sequence 103, App
c 870	6	0.5	111	10	US-09-764-870-400	Sequence 400, App	c 942	6	0.5	125	10	US-09-853-161-103	Sequence 103, App
c 871	6	0.5	111	10	US-09-925-301-1586	Sequence 1586, Ap	c 943	6	0.5	125	10	US-09-852-659A-103	Sequence 366, App
c 872	6	0.5	111	10	US-09-764-847-761	Sequence 761, App	c 944	6	0.5	125	10	US-09-867-550-366	Sequence 78, Appl
c 873	6	0.5	111	10	US-09-263-959-349	Sequence 349, App	c 945	6	0.5	125	10	US-09-965-099-78	Sequence 6, Appli
c 874	6	0.5	111	12	US-10-006-867-130	Sequence 130, App	c 946	6	0.5	126	9	US-10-051-852-78	Sequence 6, Appli
c 875	6	0.5	112	12	US-10-052-586-382	Sequence 382, App	c 947	6	0.5	126	9	US-09-813-398-6	Sequence 6, Appli
c 876	6	0.5	112	10	US-09-864-761-35730	Sequence 35730, A	c 948	6	0.5	126	10	US-09-760-723-6	Sequence 6, Appli
c 877	6	0.5	113	10	US-09-934-054-8	Sequence 8, Appli	c 949	6	0.5	126	10	US-09-355-925-6	Sequence 33598, A
c 878	6	0.5	113	10	US-09-159-469-67	Sequence 67, Appl	c 950	6	0.5	126	10	US-09-864-761-33598	Sequence 10722, A
c 879	6	0.5	113	10	US-09-798-042-67	Sequence 67, Appl	c 951	6	0.5	126	10	US-09-815-242-10722	Sequence 13036, A
c 880	6	0.5	113	10	US-09-815-248-2	Sequence 2, Appli	c 952	6	0.5	126	10	US-09-815-242-13036	Sequence 86, Appl
					Sequence 194, App	c 953							

; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 817
; LENGTH: 159

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (121)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (139)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-817

Alignment Scores:
Pred. No.: 2,73e-96 Length: 159
Score: 113.00 Matches: 113
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.59% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-764-853-817 (1-159)

QY 3424 GGCACCTAATAGTCATCGCTACCAACAAACCGGCGAGAATTCCTTTAGCTGGAGCTC 3483
Db 8 GlyThryTyrLysCysIlealThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeu 27

QY 3484 TCTGTAGTACCCAAAGAGGTGAGAAAGCACCTGTGATCTGTGAGAAACTACAGAACTGC 3543
Db 28 SerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys 47

QY 3544 GGTGTTCCGAGGCCACCCGTCGAGACTGGAGTCGGCGGTGATAGCATGCCCCACCT 3603
Db 48 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProPro 67

QY 3604 GTGTTCTACTGGAAGAAACATGAGACCATCCCTTGACACAGAGAGAGGATCAGTATG 3663
Db 68 ValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet 87

QY 3664 CACCAGACACACAGGGGTATGCTCCCTTCTCATTTACAGCCAGCAAGAAATCAGACGCT 3723
Db 88 HisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAla 107

QY 3724 GGATGTACACGTGTGCACCCAGATGAGCCGGCATC 3762
Db 108 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle 120

RESULT 3
US-09-925-299-1004
; Sequence 1004, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1004

; LENGTH: 544
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (531)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1004

Alignment Scores:
Pred. No.: 2,52e-09 Length: 544
Score: 19,00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-925-299-1004 (1-544)

QY 3034 GATGACGATGGCAACTACACCATCATGGCAGCAACCCCGGGGAGATCAGCTGT 3090
Db 229 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 247

RESULT 4

US-09-864-761-45408
; Sequence 45408, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45408
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL137225.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.75
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
US-09-864-761-45408

Alignment Scores:
Pred. No.: 57.4 Length: 17
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990b-1 (1-3963) x US-09-864-761-45408 (1-17)

QY 1185 TCCAGCAGTACCCAGCCAGCA 1208
Db 2 SerSerThrProSerProAla 9

RESULT 5
US-09-864-761-38762
; Sequence 38762, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomlca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38762
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002472.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: SWISSPROT HIT: P03179, EVALUATE 4.80e+00
US-09-864-761-38762

Alignment Scores:
Pred. No.: 51.2 Length: 46
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990b-1 (1-3963) x US-09-864-761-38762 (1-46)

QY 3146 CCAGCAGAGGTAGCCGACTGCGA 3123
Db 35 ProAlaGluValSerArgLeuArg 42

RESULT 6
US-09-864-761-39357
; Sequence 39357, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomlca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39357
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004590.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: SWISSPROT HIT: Q15438, EVALUE 4.00e-10
; OTHER INFORMATION: EST_HUMAN HIT: BF690441.1, EVALUE 2.90e+00
US-09-864-761-39357

```

```

Alignment Scores:
Pred. No.: 50.6 Length: 51
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

```

US-09-818-990B-1 (1-3963) x US-09-864-761-39357 (1-51)

```

QY 196 TTTCTGAGCCCAAGAAGATTAGAC 219
Db 11 PheLeuSerGlnGluLeuAsp 18

```

```

RESULT 7
US-09-867-550-1662
; Sequence 1662, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1662

```

```

; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-1662
Alignment Scores:
Pred. No.: 48.5 Length: 73
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

```

US-09-818-990B-1 (1-3963) x US-09-867-550-1662 (1-73)

```

QY 111 CTCCTCAACCCCTTGCCATTCGG 134
Db 2 LeuLeuGlnProLeuProPheArg 9

```

RESULT 8

```

US-09-858-664A-23
; Sequence 23, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-23

```

```

Alignment Scores:
Pred. No.: 47.1 Length: 94
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

```

US-09-818-990B-1 (1-3963) x US-09-858-664A-23 (1-94)

```

QY 3313 ACATGGCTACTCAATGGCCAACT 3336
Db 34 ThrTriLeuLeuAsnGlyGlnPro 41

```

RESULT 9

```

US-09-858-664A-30
; Sequence 30, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 119

```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-30

Alignment Scores:
Pred. No.: 45.8 Length: 119
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-858-664A-30 (1-119)
QY 3313 ACATGGCTACTCAATGGCCCAACCT 3336
|||||
Db 57 ThrTrpLeuLeuAsnGlyGlnPro 64

RESULT 10
US-09-925-302-757
; Sequence 757, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 757
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (210)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-757

Alignment Scores:
Pred. No.: 41.9 Length: 257
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-925-302-757 (1-257)
QY 174 ATCTGGCCTCCGCTCCTTACG 151
|||||
Db 90 IleLeuAlaSerAlaSerPheSer 97

RESULT 11
US-09-925-297-706
; Sequence 706, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 706
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (173)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-706

Alignment Scores:
Pred. No.: 40.6 Length: 339
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-925-297-706 (1-339)
QY 196 TTTCTGAGCCCAAGAATAATTAGAC 219
|||||
Db 310 PheLeuSerGlnGluLeuAsp 317

RESULT 12
US-09-134-262-2
; Sequence 2, Application US/09134262
; Patent No. US20010002489A1
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content in Soybean Oil
; FILE REFERENCE: Docket #17030
; CURRENT APPLICATION NUMBER: US/09/134,262
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: IBM PC; Windows NT 4.0; Microsoft Word For Windows 7.0a
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-134-262-2

Alignment Scores:
Pred. No.: 40.4 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-134-262-2 (1-352)
QY 1382 CTCATCTCCTAAGGTTGAGTGT 1405
|||||
Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 13
US-09-801-368-330
; Sequence 330, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
```

```

; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 330
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-330

Alignment Scores:
Pred. No.: 40.1 Length: 378
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-801-368-330 (1-378)
QY 2795 ATTTCATCATCTGATTCATCAACA 2772
|||||
Db 118 lIeSerSerAspSerSerThr 125

RESULT 14
US-09-815-242-12107
; Sequence 12107, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlisen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12107
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12107

Alignment Scores:
Pred. No.: 38 Length: 595
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-815-242-12107 (1-595)
QY 2371 CTGTTGGCCTGGTGGAGTGGCT 2348
|||||
Db 174 LeuLeuGlyLeuValGlyValAla 181

RESULT 15
US-09-777-710A-1
; Sequence 1, Application US/09777710A
; Patent No. US20020058305A1
; GENERAL INFORMATION:
; APPLICANT: OKINO, No. US20020058305A1omu et al.
; TITLE OF INVENTION: CERAMIDASE GENE
; FILE REFERENCE: 1422-0458P
; CURRENT APPLICATION NUMBER: US/09/777,710A
; CURRENT FILING DATE: 2001-02-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-1

Alignment Scores:
Pred. No.: 37.6 Length: 646
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-777-710A-1 (1-646)
QY 718 CAGGCTGCCAGTGAGCGGCTGGT 741
|||||
Db 425 GluAlaAlaSerGluAlaAlaGly 432

Search completed: November 30, 2002, 19:03:49
Job time : 38.5 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:38:34 ; Search time 52.5 Seconds
(without alignments)
14513.551 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacgacagcataga.....tggagagtgatgaactttaa 3963

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODE=frame_n2p.model -DEV-xmlh
-Q/cgn2_1/USPTO_spool/US09818990/runat_26112002_093222_22067/app_query.fasta_1.4103
-DB=PIR_73 -QFMT=fastan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOPCFI=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000
-DOCALIGN=200 -THR.SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990_@CGN_1.1.65_@runat_26112002_093222_22067 -NCPU=6 -ICPU=3
-NO_XLIPXY -NO_MMAL -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	19	1.4 772 2 T13078	KIAA0992 protein -
C 2	9	0.7 156 2 G72755	hypothetical prote
3	9	0.7 393 1 HVRKC1	Ig mu chain C regi
4	9	0.7 438 1 HVRKC2	Ig mu chain C regi
5	9	0.7 438 1 HVRKCS	Ig mu chain C regi
6	9	0.7 461 1 HVRKC0	Ig mu chain C regi
C 7	9	0.7 490 2 C84091	hypothetical prote
8	9	0.7 587 2 JC5669	Ca2+/calmodulin-de
C 9	9	0.7 897 2 S67283	hypothetical prote
10	9	0.7 1756 2 S45867	Tys protein - yeas
11	8	0.6 102 2 F83374	hypothetical prote
C 12	8	0.6 107 2 T44493	conserved hypotet
13	8	0.6 109 2 T29794	hypothetical prote
14	8	0.6 122 2 B82271	hypothetical prote

122	0.6	8	C 15	D71106	hypothetical prote
130	0.6	8	C 16	WMBEH3	UL33 protein - hum
141	0.6	8	C 17	S15785	heat-stable antige
164	0.6	8	C 18	S37166	pathogenesis-relat
192	0.6	8	C 19	H91103	probable transport
192	0.6	8	C 20	C85949	hypothetical prote
192	0.6	8	C 21	AG0875	probable membrane
198	0.6	8	C 22	AF1318	hypothetical prote
198	0.6	8	C 23	AF1690	hypothetical prote
199	0.6	8	C 24	E85537	hypothetical prote
199	0.6	8	C 25	A90687	hypothetical prote
204	0.6	8	C 26	S60325	transcription fact
204	0.6	8	C 27	F87295	conserved hypotet
228	0.6	8	C 28	F90130	hypothetical prote
239	0.6	8	C 29	B82361	conserved hypotet
260	0.6	8	C 30	A75572	conserved hypotet
272	0.6	8	C 31	S24375	hypothetical prote
273	0.6	8	C 32	A11989	hypothetical prote
274	0.6	8	C 33	B70695	probable enoyl-coA
295	0.6	8	C 34	AC1268	cell-shape determi
302	0.6	8	C 35	A47126	alsSD operon activ
312	0.6	8	C 36	B70896	probable coaA prot
313	0.6	8	C 37	B96692	hypothetical prote
322	0.6	8	C 38	G72421	iron(III) ABC tran
323	0.6	8	C 39	S16318	homeotic protein H
356	0.6	8	C 40	E84199	hypothetical prote
360	0.6	8	C 41	H95172	hypothetical prote
360	0.6	8	C 42	G98038	DEAD RNA helicase
362	0.6	8	C 43	S69197	oleoyl-[acyl-carri
362	0.6	8	C 44	T14393	probable oleoyl-la
369	0.6	8	C 45	S28950	3-methyl-2-oxobuta
369	0.6	8	C 46	D81261	probable ABC trans
374	0.6	8	C 47	JC1313	pectate lyase (EC
374	0.6	8	C 48	S49306	pectate lyase 2 pr
378	0.6	8	C 49	S61992	SLG1 protein - yea
381	0.6	8	C 50	C90543	hypothetical prote
382	0.6	8	C 51	S40987	hypothetical prote
383	0.6	8	C 52	AF3503	benzoate membrane
390	0.6	8	C 53	S39807	3-methyl-2-oxobuta
390	0.6	8	C 54	S75715	molybdopterin bios
392	0.6	8	C 55	A34267	3-methyl-2-oxobuta
392	0.6	8	C 56	A37157	3-methyl-2-oxobuta
401	0.6	8	C 57	AI2842	carbamoylphosphate
415	0.6	8	C 58	F84393	threonine synthase
424	0.6	8	C 59	G81160	uroporphyrin-III C
424	0.6	8	C 60	D81946	hypothetical prote
427	0.6	8	C 61	AC0203	chemotaxis MotB pr
428	0.6	8	C 62	AC2986	sugar kinase (impo
456	0.6	8	C 63	B97620	hypothetical prote
460	0.6	8	C 64	D98297	probable rhamnose
463	0.6	8	C 65	T38111	atrazine chlorohyd
478	0.6	8	C 66	A82336	probable transport
487	0.6	8	C 67	T34858	catalase (EC 1.1.1.
504	0.6	8	C 68	I53868	alpha-interneixin -
511	0.6	8	C 69	H84536	probable sugar tra
511	0.6	8	C 70	A84537	probable sugar tra
516	0.6	8	C 71	JS0083	cellulose 1,4-beta
516	0.6	8	C 72	S33164	cellulose 1,4-beta
519	0.6	8	C 73	G84598	probable bZIP tran
528	0.6	8	C 74	B84743	hypothetical prote
531	0.6	8	C 75	T02264	hypothetical prote
557	0.6	8	C 76	S62002	hypothetical prote
559	0.6	8	C 77	AI2227	hypothetical prote
595	0.6	8	C 78	C82960	probable two-compo
598	0.6	8	C 79	C96756	receptor-like prot
624	0.6	8	C 80	S67382	hypothetical prote
628	0.6	8	C 81	AG3357	peptidylprolyl iso
630	0.6	8	C 82	T38023	conserved hypotet
670	0.6	8	C 83	C83540	glycyl-tRNA synthet
684	0.6	8	C 84	T83645	hypothetical prote
694	0.6	8	C 85	T10565	hypothetical prote
717	0.6	8	C 86	F82613	VacB protein Xf198
717	0.6	8	C 87	T49238	hypothetical prote

88	8	0.6	748	2	T04011	hypothetical prote	161	7	0.5	120	2	S50867	4E-BP2 protein - h
89	8	0.6	767	2	T19690	hypothetical prote	c 162	7	0.5	121	2	PT0378	Ig heavy chain v r
90	8	0.6	834	2	R82673	hypothetical prote	c 163	7	0.5	121	2	PT0391	Ig heavy chain v r
91	8	0.6	881	2	T52601	squamosa promoter	c 164	7	0.5	121	2	E30502	Ig heavy chain v r
92	8	0.6	881	2	T52602	squamosa promoter	c 165	7	0.5	121	2	E64391	hypothetical prote
93	8	0.6	925	2	D59105	hypothetical prote	c 166	7	0.5	121	2	G90435	hypothetical prote
94	8	0.6	1051	2	A35761	cell surface glyco	c 167	7	0.5	122	1	AVMSX2	Ig heavy chain v r
95	8	0.6	1095	1	A31225	phospholipase C (E	c 168	7	0.5	122	2	AI3651	hypothetical prote
96	8	0.6	1152	2	H86486	protein Tyl/copa-	c 169	7	0.5	123	2	PT0380	Ig heavy chain v r
97	8	0.6	1337	1	I38670	protein-tyrosine-p	c 170	7	0.5	123	2	PT0381	Ig heavy chain v r
98	8	0.6	1339	2	A84683	probable SNF2 subf	c 171	7	0.5	123	2	PT0383	Ig heavy chain v r
99	8	0.6	1527	2	JE0336	canalicular multis	c 172	7	0.5	123	2	PT0384	Ig heavy chain v r
100	8	0.6	1581	2	B71636	hypothetical prote	c 173	7	0.5	123	2	PT0385	Ig heavy chain v r
101	8	0.6	1583	2	F97846	hypothetical prote	c 174	7	0.5	123	2	PT0386	Ig heavy chain v r
102	8	0.6	1597	2	S68420	citron - mouse	c 175	7	0.5	123	2	T45375	hypothetical prote
103	8	0.6	1894	2	JC4980	plexin 1 precursor	c 176	7	0.5	124	1	E69191	conserved hypotet
104	8	0.6	1905	2	T51553	Plexin - African c	c 177	7	0.5	124	2	S25113	insulin-like growt
105	8	0.6	2023	2	T13154	polycomb protein e	c 178	7	0.5	124	2	PT0388	Ig heavy chain v r
106	8	0.6	2102	2	T15626	hypothetical prote	c 179	7	0.5	124	2	PT0389	Ig heavy chain v r
107	8	0.6	2559	2	T30850	fat facets protein	c 180	7	0.5	124	2	F84561	hypothetical prote
108	8	0.6	2747	2	B49132	fat facets (faf) s	c 181	7	0.5	125	2	A53692	synapsin I - mouse
109	8	0.6	4924	2	T50176	probable peptide s	c 182	7	0.5	125	2	E72716	hypothetical prote
110	7	0.5	41	2	A44536	T-cell receptor al	c 183	7	0.5	126	2	S35887	AL5 protein - indl
111	7	0.5	43	2	I41026	colicin I0 lysis p	c 184	7	0.5	126	2	AD2523	hypothetical prote
112	7	0.5	45	2	S10921	colicin E1 lysis p	c 185	7	0.5	126	2	AI2510	hypothetical prote
113	7	0.5	47	2	FC4178	DNA-directed DNA p	c 186	7	0.5	130	2	E86906	ribosomal protein
114	7	0.5	49	1	ZHECP3	colicin lysis prot	c 187	7	0.5	131	2	B60725	hypothetical prote
115	7	0.5	49	2	H97535	colicin N lysis pr	c 188	7	0.5	134	2	H75316	conserved hypotet
116	7	0.5	52	1	ZHECN4	interleukin-5 rece	c 189	7	0.5	135	2	T32385	repressor protein
117	7	0.5	63	2	A60235	pepi protein - sta	c 190	7	0.5	135	2	T49640	hypothetical prote
118	7	0.5	69	2	S8357	hypothetical prote	c 191	7	0.5	136	2	H81891	hypothetical prote
119	7	0.5	71	2	B81812	photosystem II pho	c 192	7	0.5	137	2	S54219	flagellar basal bo
120	7	0.5	73	1	T2NT0P	hypothetical prote	c 193	7	0.5	137	2	AD0219	flagellar basal-bo
121	7	0.5	76	2	T29579	GASr1-like protein	c 194	7	0.5	137	2	S28247	NADH2 dehydrogenas
122	7	0.5	80	2	H96775	hypothetical prote	c 195	7	0.5	138	2	E75509	hypothetical prote
123	7	0.5	80	2	A81361	Ig heavy chain v r	c 196	7	0.5	138	2	C87376	beta/gamma crystal
124	7	0.5	91	2	S24523	hypothetical prote	c 197	7	0.5	138	2	H97488	hypothetical prote
125	7	0.5	92	2	C81022	hypothetical prote	c 198	7	0.5	138	2	AH2706	ADP-ribose Pyropho
126	7	0.5	93	2	H36808	antifungal protein	c 199	7	0.5	139	2	B86144	P6F3.17 protein -
127	7	0.5	94	2	S44064	conserved membrane	c 200	7	0.5	139	2	A36713	proteinase umub (E
128	7	0.5	95	2	B87216	hypothetical prote	c 201	7	0.5	139	2	AF0755	hypothetical prote
129	7	0.5	95	2	E81204	hypothetical prote	c 202	7	0.5	142	2	AG2994	transporter, dne f
130	7	0.5	100	2	F81966	hypothetical prote	c 203	7	0.5	143	2	AC3339	subtilisin inhibit
131	7	0.5	100	2	S67141	hypothetical prote	c 204	7	0.5	144	1	XSSNA	hypothetical prote
132	7	0.5	100	2	H72680	Ig heavy chain v r	c 205	7	0.5	144	2	C71252	Ig heavy chain pre
133	7	0.5	101	2	PU0002	hypothetical prote	c 206	7	0.5	145	2	S03844	conserved hypotet
134	7	0.5	102	2	H75352	hypothetical prote	c 207	7	0.5	146	2	C69983	conserved hypotet
135	7	0.5	102	2	T45500	Ig heavy chain v r	c 208	7	0.5	146	2	T14667	probable prolamin
136	7	0.5	105	2	PL0255	hypothetical prote	c 209	7	0.5	146	2	AI2918	hypothetical prote
137	7	0.5	106	2	JQ1431	conserved hypotet	c 210	7	0.5	147	2	T16672	hypothetical prote
138	7	0.5	107	2	B81031	hypothetical prote	c 211	7	0.5	148	2	T02665	hypothetical prote
139	7	0.5	109	2	A29864	somatotropin, vari	c 212	7	0.5	150	2	H97752	hypothetical prote
140	7	0.5	110	2	C71160	hypothetical prote	c 213	7	0.5	150	2	S75993	hypothetical prote
141	7	0.5	111	2	PH1028	Ig heavy chain v r	c 214	7	0.5	151	2	G86760	diacylglycerol kin
142	7	0.5	111	2	B75351	conserved hypotet	c 215	7	0.5	151	2	T14948	hypothetical prote
143	7	0.5	113	2	PH1029	Ig heavy chain v r	c 216	7	0.5	151	2	E71547	hypothetical prote
144	7	0.5	113	2	S06684	photosystem I 8.4K	c 217	7	0.5	152	2	A87261	hypothetical prote
145	7	0.5	114	2	PH1027	Ig heavy chain v r	c 218	7	0.5	152	2	T00877	hypothetical prote
146	7	0.5	114	2	I38224	protein-serine/thr	c 219	7	0.5	154	2	C98289	hypothetical prote
147	7	0.5	114	2	T25909	hypothetical prote	c 220	7	0.5	154	2	S73656	MG288 homolog F04
148	7	0.5	114	2	H89937	hypothetical prote	c 221	7	0.5	154	2	E72485	hypothetical prote
149	7	0.5	115	2	S78270	ribosomal protein	c 222	7	0.5	154	2	E72485	E6 protein - human
150	7	0.5	116	2	I51338	Ig lambda chain C	c 223	7	0.5	156	1	W6WL41	hypothetical prote
151	7	0.5	117	2	F81025	hypothetical prote	c 224	7	0.5	156	2	A87366	hypothetical prote
152	7	0.5	117	2	D81969	probable Arp synth	c 225	7	0.5	156	2	S52246	transposable retro
153	7	0.5	117	2	H95130	transcription regu	c 226	7	0.5	157	2	C81443	molybdenum cofacto
154	7	0.5	118	2	S24527	Ig heavy chain v r	c 227	7	0.5	157	2	C82651	hypothetical prote
155	7	0.5	118	2	A32673	nicotinic acetylch	c 228	7	0.5	159	2	A36107	ribosomal protein
156	7	0.5	118	2	T07177	probable oleoyl-fa	c 229	7	0.5	159	2	H86174	hypothetical prote
157	7	0.5	118	2	S27476	hypothetical prote	c 230	7	0.5	159	2	A90292	hypothetical prote
158	7	0.5	119	2	S24522	Ig heavy chain v r	c 231	7	0.5	160	2	G71122	hypothetical prote
159	7	0.5	119	2	S24497	Ig heavy chain v r	c 232	7	0.5	160	2	G90367	hypothetical prote
160	7	0.5	120	2	PT0393	Ig heavy chain v r	c 233	7	0.5	160	2	E82441	conserved hypotet

c 234	7	0.5	161	2	H69184	conserved hypothet	c 307	7	0.5	209	2	H69037	conserved hypothet
c 235	7	0.5	161	2	F82637	conserved hypothet	308	7	0.5	210	2	C84278	hypothetical prote
c 236	7	0.5	163	2	S33418	complement C4 prot	309	7	0.5	211	2	H72608	hypothetical prote
c 237	7	0.5	163	2	F72762	hypothetical prote	c 310	7	0.5	212	2	T43308	transaldolase (EC
c 238	7	0.5	164	2	A86023	hypothetical prote	c 311	7	0.5	214	2	G64432	acetyltransferase
c 239	7	0.5	164	2	H91176	ShuX-like protein	312	7	0.5	214	2	T10681	hypothetical prote
c 240	7	0.5	165	2	T45271	probable mini-circ	313	7	0.5	215	2	H75282	conserved hypothet
c 241	7	0.5	165	2	A97105	hypothetical prote	314	7	0.5	217	2	T36951	conserved hypothet
c 242	7	0.5	166	2	H87652	chemotaxis protein	315	7	0.5	217	2	T40730	probable rna-bindi
c 243	7	0.5	167	2	B98088	conserved hypothet	316	7	0.5	218	2	C71031	probable ribosomal
c 244	7	0.5	168	2	AF2820	conserved hypothet	317	7	0.5	218	2	G70654	hypothetical prote
c 245	7	0.5	171	2	E97386	hypothetical prote	318	7	0.5	219	2	S71472	endo-1,4-beta-xyla
c 246	7	0.5	171	2	AE2604	conserved hypothet	319	7	0.5	219	2	T38809	probable lipote-p
c 247	7	0.5	174	2	E69038	heat shock protein	320	7	0.5	220	2	F64203	glycerol uptake fa
c 248	7	0.5	174	2	T46382	hypothetical prote	c 321	7	0.5	221	2	T36767	probable two-compo
c 249	7	0.5	175	2	T45369	ribosomal protein	c 322	7	0.5	221	2	AH2510	hypothetical prote
c 250	7	0.5	175	2	T28050	hypothetical prote	c 323	7	0.5	221	2	C34768	ORF2 protein - Orf
c 251	7	0.5	176	2	D87299	acetyltransferase,	c 324	7	0.5	222	2	G83125	conserved hypothet
c 252	7	0.5	176	2	C55208	socA3 protein - My	c 325	7	0.5	223	2	I59173	glutamate decarbox
c 253	7	0.5	177	1	S74917	adenyl-yl-sulfate k	326	7	0.5	223	2	A36462	probable transmemb
c 254	7	0.5	177	2	S54778	NR-13 protein - qu	c 327	7	0.5	224	2	H97711	hypothetical prote
c 255	7	0.5	177	2	F72715	hypothetical prote	c 328	7	0.5	224	2	A26569	exotoxin A regulat
c 256	7	0.5	178	2	F82952	ATP synthase delta	c 329	7	0.5	226	2	T11485	H+-transporting tw
c 257	7	0.5	178	2	T00644	hypothetical prote	c 330	7	0.5	226	2	G83336	probable two-compo
c 258	7	0.5	178	2	E75340	conserved hypothet	331	7	0.5	227	2	H86940	probable membrane
c 259	7	0.5	178	2	S31955	penicillin-binding	c 332	7	0.5	228	1	LCMS	prolactin precursu
c 260	7	0.5	178	2	AH2592	conserved hypothet	333	7	0.5	228	2	JC6320	superoxide dismuta
c 261	7	0.5	179	2	G90003	ATP synthase delta	334	7	0.5	228	2	C75463	hypothetical prote
c 262	7	0.5	180	2	F97402	hypothetical prote	c 335	7	0.5	228	2	G86882	hypothetical prote
c 263	7	0.5	180	2	F70509	hypothetical prote	c 336	7	0.5	229	1	G64371	conserved hypothet
c 264	7	0.5	180	2	T20378	hypothetical prote	c 337	7	0.5	229	2	AD2462	hypothetical prote
c 265	7	0.5	180	2	C88465	protein B0244.9 [1	c 338	7	0.5	230	2	T33622	hypothetical prote
c 266	7	0.5	180	2	B86796	hypothetical prote	c 339	7	0.5	231	2	A42471	uroporphyrin-III C
c 267	7	0.5	180	2	H82480	probable acetyltra	c 340	7	0.5	232	1	JH0597	transcription fact
c 268	7	0.5	180	2	B97375	hypothetical 25.1k	c 341	7	0.5	232	2	A83396	probable COA trans
c 269	7	0.5	181	2	T48558	hypothetical prote	c 342	7	0.5	232	2	T21526	hypothetical prote
c 270	7	0.5	182	2	D83530	cytochrome b561 PA	c 343	7	0.5	232	2	T31524	hypothetical prote
c 271	7	0.5	182	2	T16472	hypothetical prote	c 344	7	0.5	233	2	T08326	hypothetical prote
c 272	7	0.5	183	2	B29017	zein zcl - maize	c 345	7	0.5	234	2	S60885	ferric exochelin u
c 273	7	0.5	185	2	C86705	hypothetical prote	c 346	7	0.5	235	2	PQ0773	4-coumarate-CoA li
c 274	7	0.5	186	2	H69090	conserved hypothet	347	7	0.5	235	2	S48924	hypothetical prote
c 275	7	0.5	186	2	G86492	polymorphic outer	348	7	0.5	235	2	D84602	hypothetical prote
c 276	7	0.5	187	2	S07271	p-aminobenzoate sy	c 349	7	0.5	235	2	A96575	hypothetical prote
c 277	7	0.5	187	2	T06581	probable deoxyribo	c 350	7	0.5	237	2	S55378	serine proteinase
c 278	7	0.5	188	2	S31952	penicillin-binding	351	7	0.5	237	2	F43692	T4 protein - rabbi
c 279	7	0.5	188	2	C87649	OmpA family protei	c 352	7	0.5	237	2	H72225	conserved hypothet
c 280	7	0.5	189	2	D69886	hypothetical prote	c 353	7	0.5	237	2	T47755	hypothetical prote
c 281	7	0.5	189	2	AD0736	probable membrane	354	7	0.5	237	2	T25152	hypothetical prote
c 282	7	0.5	190	2	A47569	pl43 DNA helicase	c 355	7	0.5	238	2	F87605	hypothetical prote
c 283	7	0.5	190	2	T10740	carbonate dehydrat	c 356	7	0.5	239	2	G97553	sugar fermentation
c 284	7	0.5	190	2	A44455	prostaglandin-D sy	c 357	7	0.5	239	2	A12773	probable amino aci
c 285	7	0.5	192	2	G83409	hypothetical prote	c 358	7	0.5	240	2	H95873	hypothetical prote
c 286	7	0.5	193	2	H82028	hypothetical prote	c 359	7	0.5	240	2	AC1921	superoxide dismuta
c 287	7	0.5	195	2	C75490	hypothetical prote	c 360	7	0.5	241	2	T47752	hypothetical prote
c 288	7	0.5	197	2	E70642	probable ribosomal	c 361	7	0.5	241	2	B69090	conserved hypothet
c 289	7	0.5	197	2	A70033	carbonic anhydrase	c 362	7	0.5	241	2	F69150	conserved hypothet
c 290	7	0.5	197	2	G81131	conserved hypothet	c 363	7	0.5	241	2	F84233	hypothetical prote
c 291	7	0.5	197	2	A86430	P26G16.12 protein	c 364	7	0.5	241	2	A87409	conserved hypothet
c 292	7	0.5	198	2	D97782	hypothetical prote	c 365	7	0.5	242	2	E88250	protein T21B10.5 l
c 293	7	0.5	198	2	S42134	light-harvesting c	c 366	7	0.5	243	2	AB1191	B. subtilis Yhf1 p
c 294	7	0.5	198	2	T33031	hypothetical prote	c 367	7	0.5	243	2	AB1549	conserved hypothet
c 295	7	0.5	200	2	C89956	30S ribosomal prot	c 368	7	0.5	244	2	AE3507	amidotransferase h
c 296	7	0.5	200	2	AD3633	hypothetical prote	c 369	7	0.5	245	2	T43036	probable transaldo
c 297	7	0.5	201	2	J01094	hypothetical 20.2k	c 370	7	0.5	245	2	T51081	hypothetical prote
c 298	7	0.5	202	2	B87438	exonuclease [impor	c 371	7	0.5	246	2	S37341	chitinase (EC 3.2.
c 299	7	0.5	203	2	G97970	hypothetical prote	c 372	7	0.5	246	2	T01710	hypothetical prote
c 300	7	0.5	204	2	AB3385	leucyl/phenylalany	c 373	7	0.5	247	2	E55545	afab protein - Esc
c 301	7	0.5	204	2	E84749	hypothetical prote	c 374	7	0.5	247	2	S45063	hypothetical prote
c 302	7	0.5	205	2	C54759	bo-type ubiquinol	c 375	7	0.5	249	2	JH0629	cleavage signal-1
c 303	7	0.5	205	2	F64935	probable CDP-alcoh	c 376	7	0.5	249	2	S03173	exotoxin A regulat
c 304	7	0.5	208	2	D85785	probable cytochrom	377	7	0.5	250	1	GWBS	indole-3-glycerol-
c 305	7	0.5	208	2	E86221	hypothetical prote	378	7	0.5	250	2	E97385	cpaD protein (AF22
c 306	7	0.5	208	2	H90936	probable cytochrom	379	7	0.5	250	2	AE2603	components of type

c 380	7	0.5	250	2	H75355	hypothetical prote	453	7	0.5	280	2	S52479	hypothetical prote
c 381	7	0.5	250	2	B83820	hypothetical prote	454	7	0.5	280	2	AD1109	B. subtilis ComBC
c 382	7	0.5	251	1	B65188	Ubiquitinone/menaqui	455	7	0.5	280	2	AF1470	C-terminal part B.
c 383	7	0.5	251	2	C98224	hypothetical prote	456	7	0.5	281	2	PC6005	methionine-tRNA li
c 384	7	0.5	251	2	A86071	hypothetical prote	c 457	7	0.5	281	2	A72561	hypothetical prote
c 385	7	0.5	251	2	B64048	hypothetical prote	458	7	0.5	282	2	H84653	probable protein k
c 386	7	0.5	253	2	AG1122	hypothetical prote	459	7	0.5	282	2	B96604	hypothetical prote
c 387	7	0.5	253	2	A11482	B. subtilis transc	460	7	0.5	282	2	A56676	homeotic protein C
c 388	7	0.5	253	2	A83719	hypothetical prote	c 461	7	0.5	282	2	S38094	hypothetical prote
c 389	7	0.5	253	2	E64420	uroporphyrin-III C	462	7	0.5	283	2	S61664	vacuolar protein v
c 390	7	0.5	253	2	T01453	hypothetical prote	c 463	7	0.5	284	2	AC2346	malate dehydrogena
c 391	7	0.5	253	2	S76719	hypothetical prote	464	7	0.5	284	2	G72107	oligopeptide trans
c 392	7	0.5	254	2	H97948	hypothetical prote	c 465	7	0.5	284	2	I51172	transcription fact
c 393	7	0.5	254	2	F95081	hypothetical prote	466	7	0.5	284	2	D83959	hypothetical prote
c 394	7	0.5	255	2	AB3579	hypothetical cytos	c 467	7	0.5	284	2	S75817	hypothetical prote
c 395	7	0.5	255	2	F83409	hypothetical prote	468	7	0.5	284	2	A96604	hypothetical prote
c 396	7	0.5	256	2	F83223	conserved hypotnet	469	7	0.5	284	2	F81563	peptide ABC transp
c 397	7	0.5	256	2	S42932	probable transmembr	470	7	0.5	284	2	A86516	oligopeptide trans
c 398	7	0.5	256	2	S42928	probable membrane-	c 471	7	0.5	285	2	S21562	hypothetical prote
c 399	7	0.5	257	2	A82691	tRNA pseudouridine	472	7	0.5	286	2	C97459	sulfate ABC transp
c 400	7	0.5	257	2	G42600	ABC-type transport	473	7	0.5	286	2	A82677	ABC transporter, m
c 401	7	0.5	257	2	AB3231	hypothetical prote	474	7	0.5	286	2	A82066	conserved hypotnet
c 402	7	0.5	257	2	C72651	hypothetical prote	c 475	7	0.5	286	2	F87075	membrane transport
c 403	7	0.5	258	2	T01481	carbonate dehydrat	c 476	7	0.5	286	2	A32349	chvD protein - Agr
c 404	7	0.5	258	2	D86371	hypothetical prote	c 477	7	0.5	287	1	B69865	probable phosphoes
c 405	7	0.5	259	1	TOH01	insulin-like growt	c 478	7	0.5	288	1	S41006	enoyl-CoA hydratase
c 406	7	0.5	259	2	T05756	hypothetical prote	479	7	0.5	288	2	G86223	hypothetical prote
c 407	7	0.5	259	2	D83557	transcription regu	480	7	0.5	288	2	T50082	azrl protein [impo
c 408	7	0.5	260	2	A87024	probable DNA-bindi	c 481	7	0.5	288	2	C70860	hypothetical prote
c 409	7	0.5	260	2	T48846	creatininase (EC 3	482	7	0.5	288	2	S58219	ABA-inducible prot
c 410	7	0.5	262	2	A10053	probable metalloen	483	7	0.5	289	2	F64374	modification methy
c 411	7	0.5	262	2	T33597	hypothetical prote	484	7	0.5	289	2	D70809	probable pabc prot
c 412	7	0.5	263	1	S23009	insulin-like growt	485	7	0.5	290	2	G75470	conserved hypotnet
c 413	7	0.5	263	2	S72528	chitinase (EC 3.2.	486	7	0.5	291	2	S27721	hypothetical prote
c 414	7	0.5	263	2	S69184	chitinase (EC 3.2.	487	7	0.5	292	2	G88448	protein C4569.4 [i
c 415	7	0.5	263	2	D82441	conserved hypotnet	488	7	0.5	292	2	S40979	hypothetical prote
c 416	7	0.5	263	2	T01614	hypothetical prote	489	7	0.5	292	2	T26879	hypothetical prote
c 417	7	0.5	264	2	S26625	chitinase (EC 3.2.	490	7	0.5	292	2	S33513	gene Fif protein -
c 418	7	0.5	264	2	D97625	hypothetical prote	491	7	0.5	292	2	F97082	transcription regu
c 419	7	0.5	264	2	AF2848	conserved hypotnet	c 492	7	0.5	293	2	S47094	hypothetical prote
c 420	7	0.5	264	2	H89005	protein T24A6.18 [493	7	0.5	294	2	C36351	sarcotoxin II-3 -
c 421	7	0.5	265	2	B83358	probable permease	494	7	0.5	294	2	A27692	sarcotoxin IIA pre
c 422	7	0.5	265	2	A36351	sarcotoxin II-1 -	c 495	7	0.5	294	2	T29028	hypothetical prote
c 423	7	0.5	266	2	C81269	probable dimethyla	496	7	0.5	294	2	E88448	protein C4569.9 [i
c 424	7	0.5	268	2	T16544	hypothetical prote	c 497	7	0.5	294	2	T36070	hypothetical prote
c 425	7	0.5	268	2	T19829	hypothetical prote	c 498	7	0.5	294	2	T33931	hypothetical prote
c 426	7	0.5	268	2	A49303	homeotic protein C	c 499	7	0.5	296	2	T05110	hypothetical prote
c 427	7	0.5	269	1	A27067	calretinin - chick	500	7	0.5	299	2	AB0440	hypothetical prote
c 428	7	0.5	271	1	A60253	calretinin - human	c 501	7	0.5	301	1	A40811	myosin-light-chain
c 429	7	0.5	271	1	S25006	calretinin - rat	502	7	0.5	301	2	B90192	conserved hypotnet
c 430	7	0.5	271	2	H84337	spermidine/putresc	503	7	0.5	302	2	A32610	spermidine synthas
c 431	7	0.5	271	2	T48994	pirin-like protein	c 504	7	0.5	302	2	A72510	hypothetical prote
c 432	7	0.5	271	2	T01767	hypothetical prote	c 505	7	0.5	303	2	C84079	lipic acid synthet
c 433	7	0.5	271	2	S62485	probable microchond	506	7	0.5	303	2	B83367	probable glycosyl
c 434	7	0.5	272	1	A36082	insulin-like growt	507	7	0.5	304	2	S04853	galactose 1-dehydr
c 435	7	0.5	272	2	I48600	insulin-like growt	508	7	0.5	305	2	H82080	UDP-3-O-3-hydroxym
c 436	7	0.5	272	2	T25044	hypothetical prote	509	7	0.5	305	2	A75211	asparaginase (EC 3
c 437	7	0.5	273	2	T22957	hypothetical prote	c 510	7	0.5	305	2	F64599	outer membrane pro
c 438	7	0.5	273	2	AF0686	hypothetical prote	c 511	7	0.5	306	1	W2WLEB	E2 protein - bovin
c 439	7	0.5	275	2	T03032	chitinase (EC 3.2.	c 512	7	0.5	306	2	C97347	oligopeptide ABC t
c 440	7	0.5	275	2	E69975	cyclodextrin metab	c 513	7	0.5	306	2	F97346	oligopeptide ABC t
c 441	7	0.5	275	2	A97226	uncharacterized co	c 514	7	0.5	306	2	C69749	conserved hypotnet
c 442	7	0.5	276	2	T08178	assemblin, striate	c 515	7	0.5	306	2	F70603	hypothetical prote
c 443	7	0.5	277	2	I37552	OX40 homolog - hum	c 516	7	0.5	306	2	G75597	hypothetical prote
c 444	7	0.5	277	2	G72357	sugar ABC transpor	c 517	7	0.5	308	2	H70767	hypothetical prote
c 445	7	0.5	278	2	AE0210	4-deoxy-L-threo-5-	518	7	0.5	309	2	T32376	hypothetical prote
c 446	7	0.5	278	2	D69959	conserved hypotnet	519	7	0.5	309	2	AI0314	probable sugar-bin
c 447	7	0.5	278	2	G87355	hypothetical prote	c 520	7	0.5	310	2	A86239	protein T10024.17
c 448	7	0.5	279	2	S76873	hypothetical prote	c 521	7	0.5	310	2	D95867	probable transcrip
c 449	7	0.5	280	2	F90845	formyltetrahydrofo	c 522	7	0.5	310	2	E64953	erfK protein precu
c 450	7	0.5	280	2	B85703	hypothetical prote	c 523	7	0.5	310	2	A99977	hypothetical prote
c 451	7	0.5	280	2	AF0649	formyltetrahydrofo	c 524	7	0.5	310	2	G85823	hypothetical prote
c 452	7	0.5	280	2	C36871	formyltetrahydrofo	c 525	7	0.5	310	2	C83088	hypothetical prote

c 526	7	0.5	311	2	JC2541	bone marrow stroma	599	7	0.5	340	2	C97653	pseudouridine synt
527	7	0.5	311	2	H84746	hypothetical prote	c 600	7	0.5	340	2	AH1986	alcohol dehydrogen
528	7	0.5	311	2	A83746	transposase (16) B	c 601	7	0.5	340	2	E95147	ribose transport s
529	7	0.5	312	2	A11871	hypothetical prote	c 602	7	0.5	341	2	A13425	F26f3.4 protein (c
c 530	7	0.5	312	2	D87475	rare lipoprotein A	c 603	7	0.5	341	2	S43586	hypothetical prote
531	7	0.5	313	2	A65140	gtmUKR operon regu	c 604	7	0.5	341	2	T04050	hypothetical prote
c 532	7	0.5	313	2	B96692	hypothetical prote	c 605	7	0.5	342	2	T29557	ferric enterobacti
c 533	7	0.5	316	2	T14643	1-aminocyclopropan	606	7	0.5	343	2	F83126	hypothetical prote
534	7	0.5	316	2	T19396	hypothetical prote	607	7	0.5	344	2	AB3236	NADH2 dehydrogenas
535	7	0.5	316	2	I38754	transcription fact	c 608	7	0.5	345	2	C45456	NADH2 dehydrogenas
c 536	7	0.5	317	2	S40535	transaldolase (EC	c 609	7	0.5	345	2	S23268	transaldolase B [i
c 537	7	0.5	317	2	H90629	transaldolase B [i	610	7	0.5	345	2	C33830	transaldolase B [i
c 538	7	0.5	317	2	H85480	transaldolase B [i	611	7	0.5	345	2	D84089	transaldolase B [i
c 539	7	0.5	317	2	AH0502	transaldolase B [i	c 612	7	0.5	345	2	AE2820	transaldolase B [i
c 540	7	0.5	317	2	AD0057	transaldolase (EC	c 613	7	0.5	346	2	B69950	transaldolase (EC
541	7	0.5	317	2	T39869	probable lysophosp	614	7	0.5	346	2	T09927	probable lysophosp
542	7	0.5	317	2	G86761	hypothetical prote	615	7	0.5	346	2	A84398	hypothetical prote
c 543	7	0.5	317	2	S72851	hypothetical prote	c 616	7	0.5	346	2	C98015	conserved hypoteth
544	7	0.5	317	2	T34228	hypothetical prote	c 617	7	0.5	346	2	E97598	fsrR protein (AF19
c 545	7	0.5	319	1	LURT5	annexin V - rat	618	7	0.5	347	2	AD2711	aldo/keto reductas
546	7	0.5	319	2	T46145	hypothetical prote	619	7	0.5	348	2	T14040	NADH2 dehydrogenas
547	7	0.5	319	2	T36845	probable membrane	c 620	7	0.5	348	2	G02297	gene N33 protein -
c 548	7	0.5	319	2	AE2823	conserved hypoteth	c 621	7	0.5	349	2	I39535	probable L-lactate
c 549	7	0.5	320	2	T42062	thioredoxin-disulf	c 622	7	0.5	349	2	G84852	hypothetical prote
c 550	7	0.5	320	2	S38670	chitinase (EC 3.2.	c 623	7	0.5	349	2	A53340	interferon regulat
551	7	0.5	321	2	G86010	regulator of glucoc	c 624	7	0.5	351	2	AI3355	UDP-3-O-[3-hydroxy
c 552	7	0.5	322	1	A53307	thioredoxin-disulf	625	7	0.5	351	2	T51513	hypothetical prote
c 553	7	0.5	322	2	T36577	thioredoxin reduct	c 626	7	0.5	352	2	T38311	protein kinase - f
c 554	7	0.5	322	2	T40834	transaldolase - fi	627	7	0.5	352	2	G00048	fusin (LESTRA) - c
555	7	0.5	323	2	QJ1552	Cl protein - panic	628	7	0.5	352	2	A45747	neuropeptide Y/pep
c 556	7	0.5	324	1	D42951	alkanal monooxygen	c 629	7	0.5	352	2	T35031	probable transcrip
557	7	0.5	325	2	B90805	probable dehydroge	630	7	0.5	352	2	F83313	conserved hypoteth
558	7	0.5	325	2	G85664	probable dehydroge	631	7	0.5	353	2	AI0220	probable sugar tra
559	7	0.5	325	2	F64845	probable 2-hydroxy	c 632	7	0.5	353	2	B72488	hypothetical prote
560	7	0.5	325	2	AI2876	pseudouridine synt	c 633	7	0.5	354	1	GNV5R	genome polyprotein
561	7	0.5	325	2	T39520	hypothetical prote	c 634	7	0.5	355	2	A42347	opsin, green-sensi
c 562	7	0.5	325	2	F90508	heat shock protein	c 635	7	0.5	355	2	F90391	conserved hypoteth
563	7	0.5	326	2	E64887	probable transposa	c 636	7	0.5	355	2	AB3516	sensory transducti
564	7	0.5	326	2	H82455	IS5 transposase VC	637	7	0.5	357	1	A54674	L-iditol 2-dehydro
c 565	7	0.5	326	2	T21924	hypothetical prote	c 638	7	0.5	360	1	A25732	inhibin alpha chai
c 566	7	0.5	328	2	G87698	cysteine synthase	c 639	7	0.5	360	2	S69063	probable membrane
c 567	7	0.5	329	2	S28305	hypothetical prote	640	7	0.5	360	2	T27324	hypothetical prote
568	7	0.5	329	2	T00873	hypothetical prote	c 641	7	0.5	361	2	T14800	hypothetical prote
569	7	0.5	329	2	AF0019	conserved hypoteth	c 642	7	0.5	361	2	A75195	d-aminopeptidase
570	7	0.5	331	2	G91164	regulator of glucoc	c 643	7	0.5	361	2	D71227	probable D-aminope
571	7	0.5	331	2	AE0481	gluconate utilizat	c 644	7	0.5	362	2	AE1397	conserved hypoteth
572	7	0.5	331	2	AB0995	gluconate utilizat	c 645	7	0.5	362	2	AH1772	conserved hypoteth
c 573	7	0.5	332	2	S63660	NADH2 dehydrogenas	c 646	7	0.5	363	2	T47588	hypothetical prote
c 574	7	0.5	332	2	AF3265	asparaginase (EC 3	c 647	7	0.5	364	1	WFFGA	hypothetical prote
c 575	7	0.5	332	2	S77370	hypothetical prote	648	7	0.5	364	2	T03892	inhibin alpha chai
c 576	7	0.5	333	2	F90325	alcohol dehydrogen	c 649	7	0.5	364	2	T25124	hypothetical prote
c 577	7	0.5	333	2	S78136	NADH2 dehydrogenas	650	7	0.5	365	2	E83952	1-deoxy-d-xylulose
c 578	7	0.5	333	2	F70672	probable thiL - My	c 651	7	0.5	365	2	E83800	hypothetical prote
579	7	0.5	333	2	T05187	chitinase (EC 3.2.	c 652	7	0.5	366	1	A24248	inhibin alpha chai
c 580	7	0.5	333	2	E86444	hypothetical prote	c 653	7	0.5	366	1	JC1106	inhibin alpha chai
581	7	0.5	333	2	G86237	protein F14N23.27	654	7	0.5	366	1	A40056	inhibin alpha chai
582	7	0.5	335	2	A72417	sugar ABC transpor	c 655	7	0.5	366	1	A40056	inhibin alpha chai
583	7	0.5	335	2	T44728	probable hydrolase	656	7	0.5	366	2	A96692	hypothetical prote
c 584	7	0.5	335	2	C65134	conserved hypoteth	657	7	0.5	367	2	E83676	pyruvate dehydroge
c 585	7	0.5	335	2	A34290	membrane protein C	658	7	0.5	367	2	S31651	MHC class I histoc
586	7	0.5	336	2	T09186	hypothetical prote	c 659	7	0.5	367	2	S25185	hypothetical prote
587	7	0.5	336	2	S61299	lipopolysaccharide	c 660	7	0.5	367	2	T00580	probable [acyl-car
588	7	0.5	336	2	B81869	lipopolysaccharide	661	7	0.5	368	2	A82249	response regulator
589	7	0.5	336	2	C81073	ADP-heptose-LPS he	c 662	7	0.5	369	2	AG3502	phenylalanine-tRNA
c 590	7	0.5	336	2	S55863	probable membrane	c 663	7	0.5	369	2	G71831	co-chaperone with
591	7	0.5	337	2	T24387	probable cysteine	c 664	7	0.5	369	2	D64686	hypothetical prote
592	7	0.5	337	2	T23794	hypothetical prote	665	7	0.5	369	2	T18663	hypothetical prote
593	7	0.5	338	1	IEEC5D	probable transposa	c 666	7	0.5	369	2	D97601	hypothetical prote
594	7	0.5	338	2	A64964	probable transposa	667	7	0.5	370	2	B97493	hypothetical prote
595	7	0.5	338	2	F70553	hypothetical prote	c 668	7	0.5	371	2	A23431	NADH2 dehydrogenas
596	7	0.5	338	2	F93035	conserved hypoteth	669	7	0.5	371	2	E97396	glutathione synthet
597	7	0.5	338	2	A89397	protein C18D4.6 [i	c 670	7	0.5	372	2	A95338	hypothetical prote
598	7	0.5	339	2	S64381	hypothetical prote	671	7	0.5	373	2	I69008	MHC class I RT1.E

c 672	7	0.5	373	2	E95102	hypothetical prote	c 745	7	0.5	409	2	D96912	o-acetylhomoserine
c 673	7	0.5	373	2	JC2426	transcription acti	c 746	7	0.5	409	2	T46741	arginine delaminase
c 674	7	0.5	374	1	A43407	cysteine synthase	c 747	7	0.5	409	2	F71828	DNA transfer prote
c 675	7	0.5	375	2	S76663	hypothetical prote	c 748	7	0.5	409	2	H82961	conserved hypothet
c 676	7	0.5	375	2	T41885	ODV-E56 orf148 - B	c 749	7	0.5	409	2	T50311	oxalisp2 protein [i
c 677	7	0.5	376	2	A55874	DNA-directed DNA p	c 750	7	0.5	409	2	T43703	respiratory protei
c 678	7	0.5	376	2	B96714	hypothetical prote	c 751	7	0.5	410	2	E86879	arginine delaminase
c 679	7	0.5	376	2	T42673	hypothetical prote	c 752	7	0.5	410	2	F82314	conserved hypothet
c 680	7	0.5	376	2	C82847	conserved hypothet	c 753	7	0.5	411	1	E71178	translation initia
c 681	7	0.5	377	2	B97376	cytochrome ba(3) (c 754	7	0.5	411	1	F75163	translation initia
c 682	7	0.5	377	2	A45851	MHC class I histoc	c 755	7	0.5	411	2	H70908	hypothetical prote
c 683	7	0.5	377	2	T04585	hypothetical prote	c 756	7	0.5	413	2	T23098	hypothetical prote
c 684	7	0.5	378	2	C64091	adenine glycosylas	c 757	7	0.5	413	2	G95305	conserved hypothet
c 685	7	0.5	379	2	E75460	conserved hypothet	c 758	7	0.5	414	2	A40350	transcription repr
c 686	7	0.5	380	2	D64129	probable 8-amino-7	c 759	7	0.5	414	2	A48273	delta/VY1/NF-EJ/UC
c 687	7	0.5	380	2	S61598	probable membrane	c 760	7	0.5	417	2	JC4698	divalent cation re
c 688	7	0.5	380	2	T04534	hypothetical prote	c 761	7	0.5	418	2	T51814	polypyrimidine tra
c 689	7	0.5	380	2	S70964	pkn5 protein - Myx	c 762	7	0.5	419	2	A25438	keratin, type I cy
c 690	7	0.5	381	1	B69030	conserved hypothet	c 763	7	0.5	419	2	F83673	PTS system, galact
c 691	7	0.5	382	1	DEKVG	glycerate dehydrog	c 764	7	0.5	420	2	A40728	microphthalimia-ass
c 692	7	0.5	382	2	B88561	protein F58A4.7b [c 765	7	0.5	420	2	JC4716	zinc finger DNA-bi
c 693	7	0.5	382	2	H75263	probable zinc prot	c 766	7	0.5	420	2	AH3015	hemolysin [importe
c 694	7	0.5	384	2	A47249	brain-specific som	c 767	7	0.5	422	1	W2WDB2	E2 protein - Bovin
c 695	7	0.5	385	2	S54992	reverse transcript	c 768	7	0.5	422	2	G88481	protein Cl6A3.9 [i
c 696	7	0.5	385	2	A86165	protein Fl3K9.6 [i	c 769	7	0.5	423	2	PQ0772	4-coumarate-CoA li
c 697	7	0.5	385	2	T18180	proline-rich prote	c 770	7	0.5	425	2	H90415	hypothetical prote
c 698	7	0.5	385	2	G89801	hypothetical prote	c 771	7	0.5	425	2	T50355	hypothetical prote
c 699	7	0.5	387	2	C83398	hypothetical prote	c 772	7	0.5	425	2	G98268	hemolysin (AF21204
c 700	7	0.5	387	2	A82849	hypothetical prote	c 773	7	0.5	425	2	S55147	hypothetical prote
c 701	7	0.5	388	2	JN0605	somatostatin recep	c 774	7	0.5	427	2	G81260	UDP-N-acetylmuramo
c 702	7	0.5	388	2	S75920	hypothetical prote	c 775	7	0.5	428	2	A43741	terminus protein -
c 703	7	0.5	389	2	T18185	cyclopropane-fatty	c 776	7	0.5	429	2	F91175	arsenical pump mem
c 704	7	0.5	389	2	H64571	cyclopropane-fatty	c 777	7	0.5	429	2	F86021	arsenical pump mem
c 705	7	0.5	389	2	B81380	hypothetical prote	c 778	7	0.5	429	2	D69635	H+/Na+-glutamate s
c 706	7	0.5	389	2	AI2230	mannose-1-phosphat	c 779	7	0.5	430	2	E64059	probable isochoris
c 707	7	0.5	389	2	AG2460	hypothetical prote	c 780	7	0.5	430	2	T05980	hypothetical prote
c 708	7	0.5	390	2	A64549	conserved hypothet	c 781	7	0.5	431	1	WJHU2G	homeotic protein H
c 709	7	0.5	390	2	E71958	hypothetical prote	c 782	7	0.5	431	2	A72294	adenylosuccinate l
c 710	7	0.5	392	1	XNHUSP	serine-pyruvate tr	c 783	7	0.5	431	2	S09824	hypothetical prote
c 711	7	0.5	392	2	S24155	alanine-glyoxylate	c 784	7	0.5	432	2	AE1170	hypothetical prote
c 712	7	0.5	393	2	A71201	hypothetical prote	c 785	7	0.5	432	2	AG1527	hypothetical prote
c 713	7	0.5	393	2	B86189	protein T25N20.9 [c 786	7	0.5	432	2	AB2222	twitching motility
c 714	7	0.5	395	2	T02185	probale translatio	c 787	7	0.5	433	2	S20963	homeotic protein H
c 715	7	0.5	395	2	T18854	hypothetical prote	c 788	7	0.5	433	2	S69999	sterigmatocystin s
c 716	7	0.5	396	2	S58161	probable translati	c 789	7	0.5	433	2	B87153	phoH-like protein
c 717	7	0.5	396	2	T04224	hypothetical prote	c 790	7	0.5	433	2	E70896	probable phoH2 pro
c 718	7	0.5	398	2	T04182	hypothetical prote	c 791	7	0.5	435	2	T00126	hypothetical prote
c 719	7	0.5	398	2	T02484	hypothetical prote	c 792	7	0.5	436	2	S47723	arsenite efflux pu
c 720	7	0.5	398	2	E87403	hypothetical prote	c 793	7	0.5	436	2	B48957	cysteine aminopept
c 721	7	0.5	398	2	C97679	bcr protein (AE006	c 794	7	0.5	436	2	F86859	aminopeptidase C [
c 722	7	0.5	398	2	AI2903	MFS permease [drug	c 795	7	0.5	436	2	JC5021	platelet-activatin
c 723	7	0.5	399	1	B49836	transcription fact	c 796	7	0.5	436	2	S64744	hypothetical prote
c 724	7	0.5	399	2	C83611	hypothetical prote	c 797	7	0.5	436	2	F71288	probable sugar ABC
c 725	7	0.5	400	1	S22539	transcription fact	c 798	7	0.5	436	2	AI2340	hypothetical prote
c 726	7	0.5	400	2	AE0198	lipoprotein releas	c 799	7	0.5	436	2	AI2340	hypothetical prote
c 727	7	0.5	401	2	F87196	probable membrane-	c 800	7	0.5	437	2	S59151	NADH2 dehydrogenas
c 728	7	0.5	401	2	T35420	probable secreted	c 801	7	0.5	437	2	T39578	nccC protein - Alc
c 729	7	0.5	401	2	T01001	hypothetical prote	c 802	7	0.5	437	2	T14192	extensin homolog T
c 730	7	0.5	401	2	AE3651	amidase (EC 3.5.1.	c 803	7	0.5	439	1	DEPSHA	homoserine dehydro
c 731	7	0.5	402	2	G83128	probable MFS trans	c 804	7	0.5	440	1	B70326	conserved hypothet
c 732	7	0.5	402	2	S32142	hypothetical prote	c 805	7	0.5	440	2	T35925	diaminopimelate de
c 733	7	0.5	403	2	AF3432	site-specific DNA-	c 806	7	0.5	441	2	E75357	probable periplasm
c 734	7	0.5	403	2	T47621	bZIP transcription	c 807	7	0.5	442	2	G71930	probable transport
c 735	7	0.5	403	2	C87315	hypothetical prote	c 808	7	0.5	442	2	B64582	sodium- and chlori
c 736	7	0.5	403	2	D87597	hypothetical prote	c 809	7	0.5	444	1	F64138	argininosuccinate
c 737	7	0.5	404	2	H87635	Rieske 2Fe-2S fami	c 810	7	0.5	444	2	B95033	aminopeptidase C [
c 738	7	0.5	404	2	H64160	hypothetical prote	c 811	7	0.5	444	2	B97904	aminopeptidase C (
c 739	7	0.5	404	2	H64175	hypothetical prote	c 812	7	0.5	444	2	T42674	hypothetical prote
c 740	7	0.5	404	2	T50335	mating pheromone r	c 813	7	0.5	445	2	JC2525	UDP-glucose dehydr
c 741	7	0.5	405	2	E69977	conserved hypothet	c 814	7	0.5	445	2	T11097	NADH2 dehydrogenas
c 742	7	0.5	406	2	A39339	protein C inhibito	c 815	7	0.5	446	2	A82337	CDP-diacylglycerol
c 743	7	0.5	406	2	T36632	probable oxidoredu	c 816	7	0.5	446	2	T25793	hypothetical prote
c 744	7	0.5	408	2	T29050	delta-1-pyrroline-	c 817	7	0.5	446	2	A38244	citrate transport

c 818	7	0.5	446	2	A42661	citrate carrier pr	c 891	7	0.5	478	1	I47154	transcription fact
c 819	7	0.5	446	2	B42661	citrate carrier pr	c 892	7	0.5	478	2	AF1758	chitinase and chit
c 820	7	0.5	446	2	AD0509	citrate-sodium sym	c 893	7	0.5	479	1	A31753	transcription fact
c 821	7	0.5	447	2	F82025	argininosuccinate	c 894	7	0.5	479	1	S22542	transcription fact
c 822	7	0.5	447	2	B81003	argininosuccinate	c 895	7	0.5	480	2	AE1121	beta-glucosidase h
c 823	7	0.5	448	2	B72262	glucose-6-phosphat	c 896	7	0.5	480	2	T46047	hypothetical prote
c 824	7	0.5	448	2	A83091	tryptophanyl-tRNA	c 897	7	0.5	481	2	E86356	hypothetical prote
c 825	7	0.5	449	1	A41386	clusterin precursor	c 898	7	0.5	481	2	A33712	metalloproteinase
c 826	7	0.5	449	2	S67819	GumC protein - Xan	c 899	7	0.5	482	2	S31478	alpha-amylase (EC
c 827	7	0.5	449	2	A84900	hypothetical prote	c 900	7	0.5	482	2	S77660	ribosomal protein
c 828	7	0.5	449	2	T50555	delta-8 sphingolip	c 901	7	0.5	482	2	AF1120	conserved hypotet
c 829	7	0.5	450	2	A54429	paired box transcr	c 902	7	0.5	485	1	S22543	transcription fact
c 830	7	0.5	450	2	T08701	hypothetical prote	c 903	7	0.5	485	1	AD0041	ramnulokinase (EC
c 831	7	0.5	452	2	F83587	inner membrane pro	c 904	7	0.5	485	2	T13694	glucuronosyltransf
c 832	7	0.5	452	2	S77436	sigma factor sibg	c 905	7	0.5	485	2	S52261	NADH2 dehydrogenas
c 833	7	0.5	452	2	T40769	hypothetical prote	c 906	7	0.5	487	1	OQECRS	vgJE protein - Esc
c 834	7	0.5	452	2	T34542	hypothetical prote	c 907	7	0.5	487	2	B91122	hypothetical prote
c 835	7	0.5	452	2	AE1710	hypothetical prote	c 908	7	0.5	487	2	A85967	hypothetical prote
c 836	7	0.5	453	2	AG1339	hypothetical prote	c 909	7	0.5	488	2	I56507	histamine H1 recep
c 837	7	0.5	453	2	F88445	protein C26E6.3 [i	c 910	7	0.5	488	2	A81006	probable malate de
c 838	7	0.5	454	2	H70158	conserved hypotet	c 911	7	0.5	488	2	B82029	probable malate de
c 839	7	0.5	455	2	AE0191	argininosuccinate	c 912	7	0.5	488	2	B95332	hypothetical prote
c 840	7	0.5	455	2	F71477	probable D-Ala/Gly	c 913	7	0.5	489	2	D87551	glutamyl-tRNA(Gln)
c 841	7	0.5	455	2	F83460	probable amino aci	c 914	7	0.5	490	1	S71776	calcium-dependent
c 842	7	0.5	455	2	JC1224	nucleobindin precu	c 915	7	0.5	490	2	PQ0164	sucrose phosphoryl
c 843	7	0.5	456	2	S30922	ferredoxin-nitrite	c 916	7	0.5	490	2	T05444	hypothetical prote
c 844	7	0.5	457	2	A56925	paired box transcr	c 917	7	0.5	491	1	I5ECKR	ketol-acid reducto
c 845	7	0.5	457	2	E88456	protein W03A5.3 [i	c 918	7	0.5	491	1	EDBEM5	immediate-early pr
c 846	7	0.5	458	2	G83690	hypothetical prote	c 919	7	0.5	491	2	D91217	ketol-acid reducto
c 847	7	0.5	459	2	T11463	NADH2 dehydrogenas	c 920	7	0.5	491	2	E86063	ketol-acid reducto
c 848	7	0.5	459	2	S52250	paired box transcr	c 921	7	0.5	491	2	AF0923	ketol-acid reducto
c 849	7	0.5	459	2	T04808	hypothetical prote	c 922	7	0.5	491	2	AH3416	AMP nucleosidase (
c 850	7	0.5	459	2	I55472	Calcium binding pr	c 923	7	0.5	493	1	S17663	NADH2 dehydrogenas
c 851	7	0.5	460	2	C27311	NADH2 dehydrogenas	c 924	7	0.5	493	2	AG2738	glutamyl-tRNA amid
c 852	7	0.5	460	2	T09956	NADH2 dehydrogenas	c 925	7	0.5	494	2	AI1480	conserved hypotet
c 853	7	0.5	460	2	T11773	NADH2 dehydrogenas	c 926	7	0.5	495	1	P1WLB	L1 protein - bovin
c 854	7	0.5	460	2	T11543	NADH2 dehydrogenas	c 927	7	0.5	496	2	T22405	protein-tyrosine k
c 855	7	0.5	460	2	T11309	NADH2 dehydrogenas	c 928	7	0.5	496	2	E97519	glutamyl-tRNA(gln)
c 856	7	0.5	460	2	S35470	NADH2 dehydrogenas	c 929	7	0.5	496	2	T41114	uroporphyrin methy
c 857	7	0.5	460	2	A58843	NADH2 dehydrogenas	c 930	7	0.5	498	2	B90456	hypothetical prote
c 858	7	0.5	460	2	B58893	NADH2 dehydrogenas	c 931	7	0.5	499	2	S70113	hypothetical prote
c 859	7	0.5	460	2	T09866	NADH2 dehydrogenas	c 932	7	0.5	501	1	P1WLB2	L1 protein - bovin
c 860	7	0.5	461	2	T19111	hypothetical prote	c 933	7	0.5	502	2	F64543	conserved hypotet
c 861	7	0.5	461	2	T11829	NADH2 dehydrogenas	c 934	7	0.5	503	2	E83642	choline sulfatase
c 862	7	0.5	461	2	S57713	probable mannosyl	c 935	7	0.5	504	2	AE3526	5-carboxymethyl-2-
c 863	7	0.5	462	2	T11136	NADH2 dehydrogenas	c 936	7	0.5	504	2	T21377	hypothetical prote
c 864	7	0.5	463	2	H97952	sodium-dependent t	c 937	7	0.5	504	2	S17248	regulatory protein
c 865	7	0.5	463	2	T25193	hypothetical prote	c 938	7	0.5	505	2	T02898	hypothetical prote
c 866	7	0.5	464	2	AF2180	diaminopimelate de	c 939	7	0.5	506	2	T19787	hypothetical prote
c 867	7	0.5	465	2	E71409	sulfate adenyllylr	c 940	7	0.5	506	2	D81703	conserved hypotet
c 868	7	0.5	465	2	D82262	probable capsular	c 941	7	0.5	507	2	S44916	2K688.6 protein -
c 869	7	0.5	465	2	T26852	hypothetical prote	c 942	7	0.5	508	2	S73540	phosphoglycerate m
c 870	7	0.5	465	2	T27885	sperm membrane pro	c 943	7	0.5	509	2	B90573	hypothetical prote
c 871	7	0.5	466	2	T30040	hypothetical prote	c 944	7	0.5	509	2	H87287	conserved hypotet
c 872	7	0.5	467	2	H69109	heat shock protein	c 945	7	0.5	511	1	E89775	2',3'-cyclic-nucle
c 873	7	0.5	468	2	H87044	probable solute-bi	c 946	7	0.5	511	2	D64905	probable sugar tra
c 874	7	0.5	468	2	B70952	probable lpgy prot	c 947	7	0.5	511	2	AC5724	hypothetical prote
c 875	7	0.5	469	2	H74657	glycosyl hydrolase	c 948	7	0.5	511	2	AC0941	probable ABC trans
c 876	7	0.5	469	2	T45201	hypothetical prote	c 949	7	0.5	511	2	H90893	hypothetical prote
c 877	7	0.5	470	2	H82054	conserved hypotet	c 950	7	0.5	511	2	T37681	kinasin-like prote
c 878	7	0.5	471	2	C71439	hypothetical prote	c 951	7	0.5	512	2	B87804	protein K04F10.2 [
c 879	7	0.5	471	2	A84741	probable myrosinas	c 952	7	0.5	512	2	PC7064	endoooligopeptidase
c 880	7	0.5	471	2	AE3384	zinc metalloprotei	c 953	7	0.5	514	1	T05655	IMP dehydrogenase
c 881	7	0.5	472	1	A49836	transcription fact	c 954	7	0.5	514	1	MMECMF	maltose transport
c 882	7	0.5	472	2	G81325	Glu-tRNAgin amidot	c 955	7	0.5	514	2	S05332	inner membrane pro
c 883	7	0.5	473	2	D83611	probable transcrip	c 956	7	0.5	514	2	S20604	transport system p
c 884	7	0.5	474	1	VYHUD	vitamin D-binding	c 957	7	0.5	514	2	H91255	part of maltose p
c 885	7	0.5	474	2	F83550	serine proteinase	c 958	7	0.5	514	2	D86096	maltose transport
c 886	7	0.5	475	2	T11586	hypothetical prote	c 959	7	0.5	514	2	AC1014	maltose transport
c 887	7	0.5	475	2	T28966	hypothetical prote	c 960	7	0.5	515	2	G83004	phosphoglycerate m
c 888	7	0.5	476	2	AC2465	6-phosphogluconate	c 961	7	0.5	515	2	S56784	hypothetical prote
c 889	7	0.5	476	2	T06792	adenylosuccinate s	c 962	7	0.5	516	2	T44002	conserved herpesvi
c 890	7	0.5	476	2	E83006	two-component resp	c 963	7	0.5	517	2	T29852	hypothetical prote

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKC1 (1-393)

QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706

DB 251 LeuProThrProLeuArgAlaSerIle 259

RESULT 4

HVRKC2

Ig mu chain C region (clone 12022) - horn shark (fragment)

C:Species: Heterodontus francisci (horn shark)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

C:Accession: S00980

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A:Title: Complete structure and organization of immunoglobulin heavy chain constant region

A:Reference number: S00980; MUID:88328985; PMID:3138109

A:Accession: S00980

A:Molecule type: mRNA

A:Residues: 1-438 <KOK>

A:Cross-references: EMBL:X07784; NID:g63963; PIDN:CAA30617.1; PID:g63964

A:Note: the sequence was determined from the differentiated gene

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger complexes.

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin

F:1-438/Domain: C region <CRE>

F:20-87/Domain: immunoglobulin homology <IM1>

F:123-190/Domain: immunoglobulin homology <IM2>

F:228-291/Domain: immunoglobulin homology <IM3>

F:330-400/Domain: immunoglobulin homology <IM4>

F:166,200,245,275,374,411,415,425/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 6.81 Length: 438
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKC2 (1-438)

QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 5

HVRKC5

Ig mu chain C region, secreted (clone 3050) - horn shark

C:Species: Heterodontus francisci (horn shark)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

C:Accession: S01853

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A:Title: Complete structure and organization of immunoglobulin heavy chain constant region

A:Reference number: S00980; MUID:88328985; PMID:3138109

A:Accession: S01853

A:Molecule type: DNA

A:Residues: 1-438 <KOK>

A:Cross-references: EMBL:X07781

A:Note: the sequence was determined from the germline gene

C:Genetics:

A:Introns: 100/1; 206/1; 309/1

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger complexes.

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin

F:1-438/Domain: C region <CRE>

F:20-87/Domain: immunoglobulin homology <IM1>

F:123-190/Domain: immunoglobulin homology <IM2>
F:228-291/Domain: immunoglobulin homology <IM3>
F:330-400/Domain: immunoglobulin homology <IM4>
F:164,200,245,275,374,411,415,425/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 6.81 Length: 438
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKCS (1-438)

QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 6

HVRKC0

Ig mu chain C region, membrane-bound (clone 3050) - horn shark

C:Species: Heterodontus francisci (horn shark)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Aug-1996

C:Accession: S01854; C32716; A46530

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A:Title: Complete structure and organization of immunoglobulin heavy chain constant region

A:Reference number: S00980; MUID:88328985; PMID:3138109

A:Accession: S01854

A:Molecule type: DNA

A:Residues: 1-461 <KOK>

A:Cross-references: EMBL:X07781

A:Note: the sequence was determined from the germline gene

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

Proc. Natl. Acad. Sci. U.S.A. 84, 5868-5872, 1987

A:Title: Extensive families of constant region genes in a phylogenetically primitive

A:Reference number: A32716; MUID:87289703; PMID:3475706

A:Accession: C32716

A:Molecule type: DNA

A:Residues: 1-99 <KO2>

A:Cross-references: GB:M17186

C:Genetics:

A:Introns: 100/1; 206/1; 309/1; 419/1; 459/3

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger complexes.

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin

F:20-87/Domain: immunoglobulin homology <IMM1>

F:123-190/Domain: immunoglobulin homology <IMM2>

F:228-291/Domain: immunoglobulin homology <IMM3>

F:330-400/Domain: immunoglobulin homology <IMM4>

F:438-458/Domain: transmembrane #status predicted <TM>

F:27-85,130-188,235-289,337-398/Disulfide bonds: #status predicted

F:164,200,245,275,374,411,415,437/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 6.78 Length: 461
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKC0 (1-461)

QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 7

CB4091

hypothetical protein BH3531 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84091
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84091
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BA07250.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3531

Alignment Scores:
Pred. No.: 6.75 Length: 490
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.70% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x C84091 (1-490)

QY 202 TCAGAAAGCTGAAGATCTGGAAGGT 176

Db 398 SerGluArgLeuLysAspLeuGluGly 406

RESULT 8

JC5669

Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.-) beta chain - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 20-Nov-1997 #sequence_revision 20-Nov-1997 #text_change 24-Sep-1999

C:Accession: JC5669; PC4493

R:Kitani, T.; Okuno, S.; Fujisawa, H.

J. Biochem. 122, 243-250, 1997

A:Title: Molecular cloning of Ca2+/calmodulin-dependent protein kinase kinase beta.

A:Reference number: JC5669; MUID:97420710; PMID:9276695

A:Accession: JC5669

A:Molecule type: mRNA

A:Residues: 1-587 <KIT>

A:Cross-references: GB:AB018081; NID:g3702720; PIDN:BAA33524.1; PID:d1034490; PID:g37027

A:Experimental source: cerebellar

A:Accession: PC4493

A:Molecule type: protein

A:Residues: 425-501 <KIT>

A:Experimental source: cerebellar

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:Keywords: ATP; phosphotransferase

F:162-445/Domain: ATP-binding #status predicted <KIN>

F:170-194/Domain: ATP-binding #status predicted <ATP>

F:480-493/Domain: calmodulin-binding #status predicted <CAB>

Alignment Scores:
Pred. No.: 6.65 Length: 587
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x JC5669 (1-587)

QY 2180 TCCCTCAGAACACACGCCAGCAA 2206

Db 571 SerProProArgThrProProGlnGln 579

RESULT 9

S67283

hypothetical protein YOR371c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O6679

C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C:Accession: S67283
R:Dalius, H.; Hebling, U.; Hofmann, B.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67261
A:Accession: S67283
A:Molecule type: DNA
A:Residues: 1-897
A:Cross-references: EMBL:Z75279; NID:gl1420801; PID:gl1420802; GSPDB:GN000015; MIPS:YOR3
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:GPE1; MIPS:YOR371c
A:Cross-references: SGD:S0005898
A:Map position: 15R

Alignment Scores:
Pred. No.: 6.44 Length: 897
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.70% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x S67283 (1-897)

QY 2795 ATTTTCATCATCTGATTCATCAACAGGA 2769

Db 93 IleSerSerAspSerSerThrGly 101

RESULT 10

S45867

TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: hypothetical protein YBR0207

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 22-Oct-1999

C:Accession: S45867

R:Entian, K.D.; Koetter, P.; Rose, M.; Li, Z.; Thermann, R.; Brendel, M.; Baur, A.; B

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45862

A:Accession: S45867

A:Molecule type: DNA

A:Residues: 1-1756 <ENT>

A:Cross-references: EMBL:Z35881; NID:g536207; PIDN:CAA84952.1; PID:g536210

A:Experimental source: strain S288C

C:Genetics:

A:Map position: 2R

A:Mobile element: retrotransposon Ty1

C:Superfamily: TyB protein

Alignment Scores:
Pred. No.: 6.12 Length: 1756
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x S45867 (1-1756)

QY 2779 GAATCAGATGATGAATTCACATGAT 2805

Db 1460 GluSerAspGluIleGlnHisAsp 1468

RESULT 11

F83374

hypothetical protein PA2174 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83374

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83374
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <STO>
A:Cross-references: GB:AE004644; GB:AE004091; NID:g9948185; PIDN:AAG05562.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2174

Alignment Scores:
Pred. No.: 86 Length: 102
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x F83374 (1-102)

QY 586 AGCTCCAGTTTCAGATCTGTCA 609
|||||
Db 70 SerSerPheSerAspLeuSer 77

RESULT 12
T44493
Conserved hypothetical protein YVAP [imported] - *Bacillus halodurans*
C:Species: *Bacillus halodurans*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Jun-2001
C:Accession: T44493; G84062
R:Takami, H.; Takaki, Y.; Nakasone, K.; Sakiyama, T.; Maeno, G.; Sasaki, R.; Hiram, C.;
Extremophiles 3, 227-233, 1999
A:Title: Genetic analysis of the chromosome of alkaliphilic *Bacillus halodurans* C-125.
A:Reference number: Z22781; MUID:99411980; PMID:10484179
A:Accession: T44493
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-107 <TAK>
A:Cross-references: EMBL:AB024562; NID:95822804; PIDN:BA83951.1; PID:95822806
A:Experimental source: strain C-125
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G84062
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07022.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3303
A:Note: yvAP
C:Superfamily: conserved hypothetical protein MTH1285

Alignment Scores:
Pred. No.: 85.6 Length: 107
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x T44493 (1-107)

QY 202 TCAGAAAGGCTGAAGATCTGGAA 179
|||||
Db 51 SerGluArgLeuIysAspLeuGlu 58

RESULT 13
T29794
hypothetical protein T14G11.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29794
R:Fulton, B.; Hawkins, J.
submitted to the EMBL data Library, November 1995
A:Description: The sequence of *C. elegans* cosmid T14G11.
A:Reference number: Z20687
A:Accession: T29794
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-109 <FUL>
A:Cross-references: EMBL:U41024; PIDN:AAA82344.1; CESP:T14G11.2
C:Genetics:
A:Gene: CESP:T14G11.2
A:Introns: 11/3; 63/2

Alignment Scores:
Pred. No.: 85.5 Length: 109
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x T29794 (1-109)

QY 3005 GCCTCTGCACATTGAATCCACTA 3028
|||||
Db 61 AlaLeuCysThrLeuAsnProLeu 68

RESULT 14
B82271
hypothetical protein VC0862 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82271
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82271
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <HEI>
A:Cross-references: GB:AE004171; GB:AE003852; NID:g9655308; PIDN:AAF94024.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0862
A:Map position: 1

Alignment Scores:
Pred. No.: 84.8 Length: 122
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x B82271 (1-122)

QY 1094 ACTCAGAAGCGCACCTAACAGG 1117
|||||
Db 102 ThrGlnLysAlaThrLeuThrArg 109

RESULT 15
D71106
hypothetical protein PH0621 - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: D71106

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 17:05:34 ; Search time 27 Seconds

(without alignments)
12175.607 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacagacagcataga.....tggagagtgaacttttaa 3963

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09818990/runat_26112002_093221_22032/app_query.fasta_1.4103
-DB=SwissProt_40 -QMT=fastcan -SUFFIX=oligo.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1_1_26 -runat_26112002_093221_22032 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	0.8	804	1 MEA6_HUMAN	O15320 homo sapien
2	9	0.7	393	1 HVC3_HETFR	P23086 heterodontu
3	9	0.7	438	1 HVC2_HETFR	P23085 heterodontu
4	9	0.7	438	1 HVC5_HETFR	P23087 heterodontu
5	9	0.7	461	1 HVCN_HETFR	P23088 heterodontu
6	8	0.6	130	1 UL33_HSV11	P10217 herpes simp
7	8	0.6	227	1 Y006_DEIRA	Q9rye3 deinococcus
8	8	0.6	302	1 ALSR_BACSU	O04778 bacillus su
9	8	0.6	312	1 CORA_WCTU	O53440 mycobacteri
10	8	0.6	323	1 HXDB_MOUSE	P23813 mus musculu
11	8	0.6	343	1 NOD1_RHISN	P55476 rhizobium s
12	8	0.6	369	1 ODBB_RAT	P35738 rattus norv
13	8	0.6	374	1 PEL2_ERWCA	P11431 erwinia car
14	8	0.6	378	1 SLG1_YEAST	P54867 saccharomyc
15	8	0.6	379	1 SGAA_METEX	P55819 methylobact
16	8	0.6	382	1 YMF4_CAEEL	P34464 caenorhabdi
17	8	0.6	390	1 MOEA_SYNY3	Q55368 synecocyst
18	8	0.6	392	1 ODBB_BOVIN	P21839 bos taurus

392	1	ODBB_HUMAN	P21953	homo sapien
401	1	CARA_AGR5	O8uf7	agrobacteri
453	1	YD14_SCHPO	Q92342	schizosacch
487	1	CATA_STRCO	Q92598	streptomyce
504	1	AINX_MOUSE	P46660	mus musculu
516	1	GUX1_PHACH	P13860	phanerochae
557	1	APG6_YEAST	Q02948	saccharomyc
603	1	REF1_DROME	Q24492	drosophila
624	1	YB3_SCHPO	Q10198	schizosacch
629	1	SKIL_MOUSE	Q60865	mus musculu
746	1	EXT1_CRIGR	O9jk82	cricetus
746	1	EXT1_HUMAN	Q16394	homo sapien
746	1	EXT1_MOUSE	P97464	mus musculu
1066	1	ITA3_CRISP	P17852	cricetidae
1095	1	PIPA_DROME	P13217	drosophila
1286	1	CTRO_HUMAN	O14578	homo sapien
1337	1	PTPJ_HUMAN	Q12913	homo sapien
1527	1	MRP3_HUMAN	O15438	homo sapien
1597	1	CTRO_MOUSE	P49025	mus musculu
1914	1	KMLS_HUMAN	O15746	homo sapien
2547	1	FAFX_HUMAN	Q93008	h probable
2559	1	FAFX_MOUSE	P70398	m probable
2747	1	FAF_DROME	P55824	drosophila
3726	1	ABF1_MOUSE	O61329	mus musculu
45	1	LYS3_SHISO	P21185	shigella so
49	1	LYS0_ECOLI	P02987	escherichia
52	1	LYS4_ECOLI	P09181	escherichia
72	1	PSBH_TOBAC	P06415	nicotiana t
87	1	Y476_MYCLE	O9cb43	mycobacteri
93	1	VG28_HSVSA	O01009	herpesvirus
94	1	AFP_ASPGI	P17737	aspergillus
113	1	PSAK_CHLRE	P14225	chlamydomon
115	1	RK22_ODOSI	P49558	odontella s
118	1	ACHE_CHICK	P16005	gallus gall
118	1	YNIF_AZOBR	P25316	azospirillu
121	1	TKNK_HUMAN	O9uhf0	homo sapien
121	1	Y733_METJA	O58143	methanococ
126	1	VAL5_ICMV	O8592	indian cass
130	1	RS9_LACIA	O9c97	lactococcus
136	1	FLGE_YEREN	Q56893	yersinia en
136	1	NB8M_BOVIN	Q02368	bos taurus
139	1	INL4_HUMAN	Q14641	homo sapien
144	1	UMUD_SALTY	P22493	salmonella
144	1	SSI_STRAO	P1006	streptomyce
144	1	YA32_TREPA	O8395	treponema p
146	1	AROF_STRLI	P55911	streptomyce
152	1	SPI8_ARATH	O64644	arabidopsis
153	1	RL22_MYCSM	O06115	mycobacteri
154	1	YF12_MYCPN	P75274	mycoplasma
156	1	VE6_HPV41	P27555	human papil
157	1	YO12_BPL2	P42547	bacterioph
159	1	RK35_SPIOL	P23326	spinacia ol
161	1	Y4LG_RHISN	P55347	rhizobium s
173	1	GLBC_NIPBR	P51536	nippostrong
174	1	GRPE_METTH	O27350	methanobact
175	1	RL22_MYCLE	O32986	mycobacteri
177	1	CYSC_SYNY3	P72940	synecocyst
177	1	NR13_COTJA	Q90343	coturnix co
183	1	ZEB2_MAZE	P08031	zea mays (m
187	1	PABA_KLEAE	P06194	klebsiella
190	1	CAH2_FLALI	P46513	flaveria li
190	1	PGHD_HUMAN	P41222	homo sapien
193	1	HUNE_DROIK	O46242	drosophila
195	1	RM32_DROME	Q9v921	drosophila
197	1	FLCP_PHATR	P95054	mycobacteri
198	1	Y660_RITCCN	Q08585	phaeodactyl
198	1	YR21_TRSRV	Q92hw0	rickettsia
201	1	YNJF_ECOLI	P25245	tomato ring
206	1	YR21_TRSRV	P76226	escherichia
214	1	YA64_METJA	Q58464	methanococ
215	1	YN68_DEIRA	Q9rrw7	deinococcus
216	1	HIS5_BRUME	Q8ye35	brucella me
216	1	HIS5_RHIME	Q92tbl	rhizobium m

92	1	7	0.5	218	RS7_PYRHO	O59230	pyrococcus	c 165	7	0.5	271	1	YAD8_SCHPO	Q09834	schizosaccc
93	1	7	0.5	219	LIPB_SCHPO	Q36017	schizosacch	c 166	7	0.5	272	1	IBP1_MOUSE	P17876	mus musculus
c 94	1	7	0.5	219	RR3_ASTJO	P58133	astasia lon	c 167	7	0.5	272	1	IBP1_RAT	P21743	rattus norv
c 95	1	7	0.5	223	ATV2A_MOUSE	P20934	mus musculus	c 168	7	0.5	276	1	KRE9_CANGA	O74683	candida gla
c 96	1	7	0.5	226	ATP6_RABIT	P079432	oryctolagus	c 169	7	0.5	276	1	SFAS_CHLRE	Q39618	chlamydomon
c 97	1	7	0.5	226	PRL_MOUSE	P06879	mus musculus	c 170	7	0.5	277	1	TNR4_HUMAN	P43489	homo sapien
c 98	1	7	0.5	228	SODM_ACITCA	Q59094	acinetobact	c 171	7	0.5	278	1	YOHM_BACSU	P34511	bacillus su
c 99	1	7	0.5	229	Y575_METJA	Q57995	methanococc	c 172	7	0.5	279	1	DAPE_SYNY3	P74667	synechocyst
c 100	1	7	0.5	230	SUMT_METIV	P29564	methanobact	c 173	7	0.5	280	1	K1C3_XENLA	P05782	xenopus lae
c 101	1	7	0.5	231	NU4M_AKAC	P92623	agkistrodon	c 174	7	0.5	280	1	PURU_ECOLI	P37051	escherichia
c 102	1	7	0.5	231	NU4M_AKGC	P92492	agkistrodon	c 175	7	0.5	280	1	PURU_SHIFL	P38480	shigella fl
c 103	1	7	0.5	231	NU4M_AKHA	Q03726	agkistrodon	c 176	7	0.5	280	1	YME1_YEAST	Q03433	saccharomyc
c 104	1	7	0.5	231	NU4M_AKGIN	Q03727	agkistrodon	c 177	7	0.5	281	1	YH71_AERPE	Q9YB24	aeropyrum p
c 105	1	7	0.5	231	NU4M_AKRP1	P92503	agkistrodon	c 178	7	0.5	282	1	CK04_MOUSE	Q07424	mus musculus
c 106	1	7	0.5	231	NU4M_AKGRH	P92613	agkistrodon	c 179	7	0.5	282	1	YD53_YEAST	P36121	saccharomyc
c 107	1	7	0.5	231	NU4M_ATRNM	O03692	atropoides	c 180	7	0.5	283	1	VAM3_YEAST	Q12241	saccharomyc
c 108	1	7	0.5	231	NU4M_ATRPI	O03695	atropoides	c 181	7	0.5	286	1	CHVD_AGRTO	P12622	agrobacteri
c 109	1	7	0.5	231	NU4M_AZFEF	P92494	azemlops fe	c 182	7	0.5	287	1	YKUE_BACSU	Q04870	bacillus su
c 110	1	7	0.5	231	NU4M_BOTAS	O03697	bothriops as	c 183	7	0.5	288	1	AZRI_SCHPO	Q09189	schizosacch
c 111	1	7	0.5	231	NU4M_BOTBI	O03698	bothriopsis	c 184	7	0.5	288	1	ECHM_CAEEL	P34559	caenorhabdi
c 112	1	7	0.5	231	NU4M_BOTER	O03699	bothriops er	c 185	7	0.5	288	1	MRGX_HUMAN	Q15014	homo sapien
c 113	1	7	0.5	231	NU4M_BOTSC	O03701	bothrieichs	c 186	7	0.5	289	1	MTM3_METJA	O58015	methanococc
c 114	1	7	0.5	231	NU4M_CAURH	O03707	causius rhom	c 187	7	0.5	291	1	YD53_SYNY3	P42350	synechocyst
c 115	1	7	0.5	231	NU4M_CERGO	O03703	cerrophadio	c 188	7	0.5	292	1	YO14_CAEEL	Q09505	caenorhabdi
c 116	1	7	0.5	231	NU4M_CROAD	O03702	crotalus ad	c 189	7	0.5	294	1	SR23_SARPE	P24490	sarcophaga
c 117	1	7	0.5	231	NU4M_CROLE	O03704	crotalus le	c 190	7	0.5	294	1	SR2_SARPE	P14667	sarcophaga
c 118	1	7	0.5	231	NU4M_CROVE	O03710	crotalus vi	c 191	7	0.5	294	1	YO19_CAEEL	Q09507	caenorhabdi
c 119	1	7	0.5	231	NU4M_HYPHY	O03733	hypnale hyp	c 192	7	0.5	295	1	KMLC_DICDI	P25323	dictyosteli
c 120	1	7	0.5	231	NU4M_LACMU	P92649	lachesis mu	c 193	7	0.5	297	1	FBRU_LEIMA	P35549	leishmania
c 121	1	7	0.5	231	NU4M_PORHY	O03763	portidium	c 194	7	0.5	302	1	SPEE_HUMAN	P19623	homo sapien
c 122	1	7	0.5	231	NU4M_PORNA	O03772	portidium	c 195	7	0.5	302	1	SPEE_MOUSE	Q64674	mus musculus
c 123	1	7	0.5	231	NU4M_PORP	O03773	portidium	c 196	7	0.5	303	1	LIPA_BACHD	Q9K7C9	bacillus ha
c 124	1	7	0.5	231	NU4M_SISMI	O03778	sistrurus m	c 197	7	0.5	304	1	GAL_PSEFL	P11866	pseudomonas
c 125	1	7	0.5	231	NU4M_TRTAB	O03780	trimeresuru	c 198	7	0.5	305	1	ASGX_PYRAB	Q9V262	pyrococcus
c 126	1	7	0.5	231	NU4M_TRICN	O03792	trimeresuru	c 199	7	0.5	305	1	LPXC_VIBCH	Q9KPH2	vibrio chol
c 127	1	7	0.5	231	NU4M_TRIEL	P92793	trimeresuru	c 200	7	0.5	306	1	ISPE_MYCTU	O05596	mycobacteri
c 128	1	7	0.5	231	NU4M_TRIFL	P92794	trimeresuru	c 201	7	0.5	306	1	YBFH_BACSU	Q31448	bacillus su
c 129	1	7	0.5	231	NU4M_TRIOK	P92681	trimeresuru	c 202	7	0.5	308	1	TATC_MYCTU	Q10702	mycobacteri
c 130	1	7	0.5	231	NU4M_TRIST	P92759	trimeresuru	c 203	7	0.5	310	1	ERFK_ECOLI	P39176	escherichia
c 131	1	7	0.5	231	NU4M_TRITWA	O03807	trimeresuru	c 204	7	0.5	310	1	TATC_MYCLE	P34078	mycobacteri
c 132	1	7	0.5	232	SLT2_CAEEL	Q19892	caenorhabdi	c 205	7	0.5	311	1	BST1_MOUSE	Q64277	mus musculus
c 133	1	7	0.5	235	RR2_GUITH	O78482	guillardia	c 206	7	0.5	316	1	TALB_ECOLI	P30148	escherichia
c 134	1	7	0.5	235	YHE4_YEAST	P38727	saccharomyc	c 207	7	0.5	316	1	TALB_SALTY	Q8XG45	salmonella
c 135	1	7	0.5	237	VT4_SFVKA	P25948	shope fibro	c 208	7	0.5	317	1	TAL_YERPE	Q84in2	versinia pe
c 136	1	7	0.5	238	SSPN_RABIT	P82352	oryctolagus	c 209	7	0.5	318	1	ANX5_RAT	P14668	rattus norv
c 137	1	7	0.5	239	SFSA_AGRF5	P58429	agrobacteri	c 210	7	0.5	321	1	TRXB_STRCL	Q05741	streptomyce
c 138	1	7	0.5	243	RNFE_RHOCA	P97055	rhodobacteri	c 211	7	0.5	321	1	TRXB_STRCO	P52215	streptomyce
c 139	1	7	0.5	246	CHID_LYCES	Q05537	lycopersico	c 212	7	0.5	322	1	TALI_SCHPO	O42700	schizosacch
c 140	1	7	0.5	247	AFAB_ECOLI	P53516	escherichia	c 213	7	0.5	323	1	VAL1_PASVK	Q00338	panicum str
c 141	1	7	0.5	247	NFAE_ECOLI	P46738	escherichia	c 214	7	0.5	324	1	LXB2_PHOLU	P23147	photorhabdu
c 142	1	7	0.5	250	TRPC_BACSU	P03964	bacillus su	c 215	7	0.5	325	1	YCDW_ECOLI	P75913	escherichia
c 143	1	7	0.5	251	UBIE_ECOLI	P27851	escherichia	c 216	7	0.5	326	1	INHS_ECOLI	P76071	escherichia
c 144	1	7	0.5	251	Y095_HAEIN	Q57060	haemophilus	c 217	7	0.5	331	1	GNTR_ECOLI	P46860	escherichia
c 145	1	7	0.5	253	YDIJ_BACHD	Q9Z9P4	bacillus ha	c 218	7	0.5	335	1	CBP3_YEAST	P21560	saccharomyc
c 146	1	7	0.5	255	GP39_BPSP1	O48395	bacterioph	c 219	7	0.5	336	1	YN66_YEAST	P42836	saccharomyc
c 147	1	7	0.5	257	NOCF_AGRF5	P35116	agrobacteri	c 220	7	0.5	337	1	BRB1_RAT	P97583	rattus norv
c 148	1	7	0.5	257	YL48_MYCLE	Q9CCE2	mycobacteri	c 221	7	0.5	338	1	INSH_ECOLI	P03837	escherichia
c 149	1	7	0.5	258	GLPE_MYCGE	P47279	mycoplasma	c 222	7	0.5	338	1	PSD7_DROME	P26270	drosophilla
c 150	1	7	0.5	259	IBP1_HUMAN	P08833	homo sapien	c 223	7	0.5	338	1	YC61_METTH	P26270	drosophilla
c 151	1	7	0.5	259	TOXR_PPSAE	P09852	pseudomonas	c 224	7	0.5	339	1	RPOA_THIBE	P92439	thiopyrum
c 152	1	7	0.5	261	DHB8_HUMAN	Q92506	homo sapien	c 225	7	0.5	339	1	TORS_DROME	O77277	drosophilla
c 153	1	7	0.5	263	IBP1_BOVIN	P24591	bos taurus	c 226	7	0.5	339	1	YG2J_YEAST	P53252	saccharomyc
c 154	1	7	0.5	264	TVSY_SALTY	Q8Zma9	salmonella	c 227	7	0.5	342	1	HXCA_HUMAN	Q9NY66	homo sapien
c 155	1	7	0.5	265	IHA_SHEEP	P38440	ovis aries	c 228	7	0.5	345	1	NQO8_PARDE	P29920	paracoccus
c 156	1	7	0.5	265	SR21_SARPE	P24491	sarcophaga	c 229	7	0.5	345	1	NUOH_RHOCA	P42032	rhodobacter
c 157	1	7	0.5	267	SSF2_HUMAN	P28290	homo sapien	c 230	7	0.5	346	1	ILVC_BUCUL	Q9a98	buchnera ap
c 158	1	7	0.5	268	CDX1_MOUSE	P18111	mus musculus	c 231	7	0.5	346	1	ILVC_BUCUM	Q9a940	buchnera ap
c 159	1	7	0.5	269	CLB2_CHICK	P07090	gallus gall	c 232	7	0.5	346	1	YOCL_BACSU	P45946	bacillus su
c 160	1	7	0.5	270	Y231_AQAE	O66421	aquifex aeo	c 233	7	0.5	348	1	N33_HUMAN	Q13454	homo sapien
c 161	1	7	0.5	271	CLB2_HUMAN	P22676	homo sapien	c 234	7	0.5	349	1	CCR4_RAT	O08565	rattus norv
c 162	1	7	0.5	271	CLB2_MOUSE	Q08331	mus musculus	c 235	7	0.5	349	1	IRF2_HUMAN	P14316	homo sapien
c 163	1	7	0.5	271	CLB2_RAT	P47728	rattus norv	c 236	7	0.5	349	1	IRF2_MOUSE	P23906	mus musculus
c 164	1	7	0.5	271	PIR4_ARATH	Q91X45	arabidopsis	c 237	7	0.5	349	1	LDH_ALCEU	Q07251	alcaligenes

c 238	7	0.5	351	1	LPXD_BRUME	Q44630 bruceella me	c 311	7	0.5	427	1	IRE3_HUMAN	Q14653 homo sapien
c 239	7	0.5	352	1	CCR4_CERTO	O62747 cercocobus	312	7	0.5	427	1	MURE_CAMJE	O69290 campylobact
c 240	7	0.5	352	1	CCR4_HUMAN	P30991 homo sapien	313	7	0.5	428	1	TERM_DROME	P11455 drosophila
c 241	7	0.5	352	1	CCR4_MACFA	Q28474 macaca fasc	314	7	0.5	429	1	ARSB_ECOLI	P37310 escherichia
c 242	7	0.5	352	1	CCR4_WACMU	P79394 macaca mula	315	7	0.5	430	1	MENF_ECOLI	P44613 haemophilus
c 243	7	0.5	352	1	CCR4_PAPAN	P56491 papio anubi	316	7	0.5	431	1	HXB3_HUMAN	P14613 homo sapien
c 244	7	0.5	353	1	PRK1_SCHPO	O13958 schizosacch	317	7	0.5	431	1	PUR8_THEMEA	Q9x010 thermotoga
c 245	7	0.5	353	1	CCR4_FELCA	P56498 felis silve	318	7	0.5	431	1	UL61_HCMVA	P16818 human cytom
c 246	7	0.5	354	1	POL1_TRSVR	P29150 tomato ring	319	7	0.5	433	1	AFUR_EMENI	P52957 emericeila
c 247	7	0.5	355	1	OPSG_CHICK	P28683 gallus gall	320	7	0.5	433	1	HXB3_MOUSE	P09026 mus musculus
c 248	7	0.5	356	1	DHSO_HUMAN	Q00796 homo sapien	321	7	0.5	435	1	PEPC_LACLC	Q04723 lactococcus
c 249	7	0.5	358	1	GP20_HUMAN	P99678 homo sapien	322	7	0.5	436	1	PAFA_CAVPO	P70683 c platelet-
c 250	7	0.5	359	1	CCR4_MOUSE	P70658 m c-x-c che	323	7	0.5	437	1	NCCC_ALCXX	Q44584 alcaligenes
c 251	7	0.5	360	1	IHA_BOVIN	P07994 bos taurus	324	7	0.5	437	1	NU4M_ALBUC	P48914 albinaria c
c 252	7	0.5	361	1	IHA_TRIVU	O77755 trichosurus	325	7	0.5	440	1	DCDA_STRCO	Q9zbt5 streptomyce
c 253	7	0.5	364	1	IHA_PIG	P04087 sus scrofa	326	7	0.5	440	1	Y284_AQUAE	O66338 aquifex aeo
c 254	7	0.5	366	1	IHA_HUMAN	P05111 homo sapien	327	7	0.5	442	1	CYS4_DICDI	P54639 dictyostell
c 255	7	0.5	366	1	IHA_MOUSE	Q04997 mus musculus	328	7	0.5	444	1	ASSY_HAEIN	P44315 haemophilus
c 256	7	0.5	366	1	IHA_RAT	P17490 rattus norv	329	7	0.5	445	1	ASSY_PASMU	P57877 pasteurella
c 257	7	0.5	366	1	IHA_RAT	P17490 rattus norv	330	7	0.5	445	1	NU4M_DINSE	O79555 dinodon sem
c 258	7	0.5	367	1	IHA_HORSE	P55101 equus cabal	331	7	0.5	445	1	YAV5_XANCV	P19520 xanthomonas
c 259	7	0.5	367	1	YMF4_STRCO	P43167 streptomyce	332	7	0.5	446	1	CITN_KLEPN	P31602 klebsiella
c 260	7	0.5	369	1	DNAJ_HELPUP	Q92192 helicobacte	333	7	0.5	446	1	CITN_SALDU	P31603 salmonella
c 261	7	0.5	369	1	DNAJ_HELPY	O25890 helicobacte	334	7	0.5	446	1	CITN_SALTY	P31604 salmonella
c 262	7	0.5	369	1	O24A_DROME	P81913 drosophila	335	7	0.5	448	1	G6PI_THEMEA	Q9xia5 thermotoga
c 263	7	0.5	371	1	NUIM_NEUCR	P08774 neurospora	336	7	0.5	449	1	CLUS_HUMAN	P10909 homo sapien
c 264	7	0.5	374	1	CYS1_CAPAN	P31300 capsicum an	337	7	0.5	450	1	PAX8_HUMAN	Q06710 homo sapien
c 265	7	0.5	375	1	DP3B_SYNPT	P52023 synechococc	338	7	0.5	456	1	TRME_BUCAP	Q44633 buchnera ap
c 266	7	0.5	375	1	OE56_NPVBM	Q92500 bombyx mori	339	7	0.5	456	1	TRME_LEPIN	P97043 leptospira
c 267	7	0.5	375	1	TRMU_CAEEL	Q17440 caenorhabdi	340	7	0.5	457	1	PAX8_MOUSE	Q00288 mus musculus
c 268	7	0.5	377	1	MTB1_BRUAB	Q03070 bruceella ab	341	7	0.5	458	1	PAX8_RAT	P51974 rattus norv
c 269	7	0.5	377	1	VE2_HPV44	Q80916 human papil	342	7	0.5	459	1	NCB1_MOUSE	Q02819 mus musculus
c 270	7	0.5	378	1	MUTY_HAEIN	P44320 haemophilus	343	7	0.5	459	1	NCB1_RAT	Q63083 rattus norv
c 271	7	0.5	380	1	BIOF_HAEIN	P44422 haemophilus	344	7	0.5	459	1	NU4M_POLOR	Q95917 polypterus
c 272	7	0.5	380	1	COSC_YEAST	P53053 saccharomyc	345	7	0.5	459	1	PAX8_CANFA	P47240 canis famil
c 273	7	0.5	380	1	PKN5_MYXXA	P54737 myxococcus	346	7	0.5	460	1	NU4M_ASTPE	P11992 asterina pe
c 274	7	0.5	380	1	YMH7_CAEEL	P34474 caenorhabdi	347	7	0.5	460	1	NU4M_BRARE	Q9mly1 brachydanio
c 275	7	0.5	382	1	DHGY_CUCSA	P13443 cucumis sat	348	7	0.5	460	1	NU4M_CARAU	Q78687 carassius a
c 276	7	0.5	382	1	DXY_BACHD	Q9ka69 bacillus ba	349	7	0.5	460	1	NU4M_CROLA	P34194 crossostoma
c 277	7	0.5	384	1	SSR4_RAT	P30937 rattus norv	350	7	0.5	460	1	NU4M_GADMO	P55781 gadus morhu
c 278	7	0.5	388	1	SSR4_HUMAN	P31391 homo sapien	351	7	0.5	460	1	NU4M_ONCMY	P11631 oncorhynchu
c 279	7	0.5	392	1	SPYA_HUMAN	P21549 homo sapien	352	7	0.5	460	1	NU4M_SALSA	Q9zxm4 salmo salar
c 280	7	0.5	392	1	SPYA_RABIT	P31030 corytolagus	353	7	0.5	460	1	NU4M_SCYCA	Q79410 scyllorhinu
c 281	7	0.5	396	1	RFIM_SCHPO	Q09691 schizosacch	354	7	0.5	460	1	NU4M_SQUAC	Q9z245 squallus aca
c 282	7	0.5	397	1	ALR1_RHIL0	Q98a05 rhizobium l	355	7	0.5	461	1	MNT2_CANAL	P46592 candida alb
c 283	7	0.5	399	1	Y507_PASMU	Q9enc5 pasteurella	356	7	0.5	461	1	NCB1_HUMAN	Q02818 homo sapien
c 284	7	0.5	401	1	LSG1_HAEIN	P71399 haemophilus	357	7	0.5	461	1	NU4M_LATCH	Q03173 latimeria c
c 285	7	0.5	402	1	YIN2_STRAM	P32426 streptomyce	358	7	0.5	465	1	OCT2_MOUSE	Q00196 mus musculus
c 286	7	0.5	404	1	Y4C1_RHISN	P55391 rhizobium s	359	7	0.5	465	1	SPE4_CAEEL	Q01608 caenorhabdi
c 287	7	0.5	404	1	Y867_HAEIN	Q57484 haemophilus	360	7	0.5	470	1	EGR2_RAT	P51774 rattus norv
c 288	7	0.5	404	1	ZFS1_SCHPO	P47979 schizosacch	361	7	0.5	471	1	PDW2_DROVI	Q24705 drosophila
c 289	7	0.5	405	1	Y57A_HUMAN	O60320 homo sapien	362	7	0.5	472	1	POU2_BRARE	Q90270 brachydanio
c 290	7	0.5	405	1	YRKO_BACSU	P54442 bacillus su	363	7	0.5	474	1	LCPT_STRIN	Q33654 streptococc
c 291	7	0.5	406	1	IPSP_HUMAN	P05154 homo sapien	364	7	0.5	474	1	VTDB_HUMAN	P02774 homo sapien
c 292	7	0.5	409	1	ARCA_LACSK	O53088 lactobacill	365	7	0.5	475	1	YDF6_SCHPO	Q10478 schizosacch
c 293	7	0.5	409	1	OX12_SCHPO	O43092 schizosacch	366	7	0.5	476	1	PURA_WHEAT	Q24396 triticum ae
c 294	7	0.5	410	1	ARCA_LACLA	P58013 lactococcus	367	7	0.5	478	1	OCT2_HUMAN	P09086 homo sapien
c 295	7	0.5	410	1	VE2_BPVI	P03122 bovine papi	368	7	0.5	478	1	OCT2_PIG	Q29013 sus scrofa
c 296	7	0.5	411	1	IF2G_PYRAB	Q9v190 pyrococcus	369	7	0.5	481	1	PRTB_ERWCH	P16316 erwinia chr
c 297	7	0.5	411	1	IF2G_PYRFO	O8u082 pyrococcus	370	7	0.5	484	1	IRX5_MOUSE	Q9jfk4 mus musculus
c 298	7	0.5	411	1	IF2G_PYRHO	O59410 pyrococcus	371	7	0.5	487	1	TTDT_ECOLI	P39414 escherichia
c 299	7	0.5	411	1	VE2_BPVI	P11299 bovine papi	372	7	0.5	488	1	HH1R_CAVPO	P31389 cavia porce
c 300	7	0.5	413	1	ARCA_BACLI	O86131 bacillus li	373	7	0.5	490	1	ILVC_ECO57	P58256 escherichia
c 301	7	0.5	413	1	ARCA_CLOPE	Q46254 clostridium	374	7	0.5	490	1	ILVC_ECOLI	P05793 escherichia
c 302	7	0.5	414	1	TYT1_HUMAN	P25490 homo sapien	375	7	0.5	490	1	ILVC_SALTY	P05989 salmonella
c 303	7	0.5	414	1	TYT1_MOUSE	Q00899 mus musculus	376	7	0.5	490	1	SUCP_LEUME	Q59495 leuconostoc
c 304	7	0.5	416	1	MTPI_PSYTA	O33481 psychrobact	377	7	0.5	491	1	VIEL_HCMVA	P13202 human cytom
c 305	7	0.5	417	1	CZCC_ALCEU	P13509 alcaligenes	378	7	0.5	493	1	NUBM_NEUCR	P24917 neurospora
c 306	7	0.5	417	1	IRX5_HUMAN	P78411 homo sapien	379	7	0.5	495	1	ENP2_RAT	Q35795 rattus norv
c 307	7	0.5	419	1	K1C4_XENLA	P05781 xenopus lae	380	7	0.5	495	1	VLL_BPVI	P03103 bovine papi
c 308	7	0.5	422	1	OAT_CAEEL	Q18040 caenorhabdi	381	7	0.5	496	1	NUBM_ASPNG	Q92406 aspergillus
c 309	7	0.5	425	1	Y450_HUMAN	O75038 homo sapien	382	7	0.5	496	1	SUMT_SCHPO	Q74468 schizosacch
c 310	7	0.5	425	1	YNN6_YEAST	P53911 saccharomyc	383	7	0.5	497	1	VLL_BPVI	P06458 bovine papi

c 384	7	0.5	497	1	WAS2_HUMAN	Q9y6w5	homo sapien	c 457	7	0.5	632	1	S6AB_HUMAN	P48066	homo sapien
c 385	7	0.5	502	1	Y190_HELPY	P56117	helicobacte	458	7	0.5	636	1	DNK2_SYNY3	P22358	synchocyst
c 386	7	0.5	504	1	MIG1_YEAST	P27705	saccharomyc	459	7	0.5	638	1	YDEC_SCHPO	Q10445	schizosacch
c 387	7	0.5	506	1	NEK3_HUMAN	P51956	homo sapien	c 460	7	0.5	640	1	SVT_THENA	Q9wzj9	thermotoga
c 388	7	0.5	507	1	ALG6_HUMAN	Q9y672	homo sapien	461	7	0.5	643	1	PGT_HUMAN	Q92959	homo sapien
c 389	7	0.5	507	1	YO26_CAEEL	P34676	caenorhabdi	462	7	0.5	643	1	PGT_RAT	Q00910	rattus norv
c 390	7	0.5	508	1	PMG1_MYCPN	P75167	mycoplasma	463	7	0.5	644	1	PAB4_HUMAN	Q13310	homo sapien
c 391	7	0.5	508	1	Y202_HUMAN	Q92599	homo sapien	c 464	7	0.5	644	1	XYND_CELFI	P54865	cellulomona
c 392	7	0.5	509	1	SYM_MYCPU	Q50319	mycoplasma	465	7	0.5	646	1	WEEL_HUMAN	P30291	homo sapien
c 393	7	0.5	510	1	ENPL_MOUSE	P55772	mus musculus	466	7	0.5	646	1	DREB_CHICK	P18302	gallus gall
c 394	7	0.5	511	1	EGO_ECOLI	P77257	escherichia	467	7	0.5	652	1	ILV3_BRANA	P27819	brassica na
c 395	7	0.5	511	1	NEK3_MOUSE	Q9r0a5	mus musculus	468	7	0.5	655	1	ILF1_HUMAN	Q01167	homo sapien
c 396	7	0.5	514	1	IMD2_MOUSE	P24547	mus musculus	469	7	0.5	655	1	ILV1_BRANA	P27818	brassica na
c 397	7	0.5	514	1	MALE_ECOLI	P02916	escherichia	c 470	7	0.5	661	1	YDJB_SCHPO	Q10369	schizosacch
c 398	7	0.5	514	1	MALE_ENTAE	P18812	enterobacte	c 471	7	0.5	662	1	NX2B_RAT	Q63376	rattus norv
c 399	7	0.5	514	1	MALE_SALTY	P26467	salmonella	c 472	7	0.5	664	1	PD13_SHEEP	Q02849	ovis aries
c 400	7	0.5	514	1	ZRF1_MOUSE	P54103	mus musculus	c 473	7	0.5	666	1	NX2B_HUMAN	P58401	homo sapien
c 401	7	0.5	515	1	IE63_HSV60	P52354	human herpe	c 474	7	0.5	667	1	OS9_HUMAN	Q13438	homo sapien
c 402	7	0.5	515	1	MAD3_YEAST	P47074	saccharomyc	475	7	0.5	667	1	Y366_MYCGE	P47606	mycoplasma
c 403	7	0.5	516	1	IE63_HSV62	P52539	human herpe	c 476	7	0.5	669	1	NAH6_HUMAN	Q92581	homo sapien
c 404	7	0.5	517	1	PLIN_RAT	P43884	rattus norv	c 477	7	0.5	670	1	OATP_HUMAN	P46731	homo sapien
c 405	7	0.5	517	1	Y652_HUMAN	Q75143	homo sapien	c 478	7	0.5	670	1	SYNL_MOUSE	Q88935	mus musculus
c 406	7	0.5	522	1	SIN1_CHICK	Q9w6s3	gallus gall	c 479	7	0.5	677	1	YI36_MYCTU	Q50597	mycobacteri
c 407	7	0.5	526	1	MITE_MOUSE	Q08874	mus musculus	c 480	7	0.5	678	1	PTR3_YEAST	P43606	saccharomyc
c 408	7	0.5	529	1	KRT1_HUMAN	O00522	homo sapien	481	7	0.5	680	1	KALM_HUMAN	P23352	homo sapien
c 409	7	0.5	530	1	TP6B_SULSO	Q97zfo	sulfolobus	c 482	7	0.5	687	1	VIUA_VIBCH	Q00964	vibrio chol
c 410	7	0.5	537	1	SY63_DISOM	P24507	discopoge o	c 483	7	0.5	687	1	WHIT_DROME	P10090	drosophila
c 411	7	0.5	542	1	CDP3_ORYSA	P53684	oryza sativ	484	7	0.5	689	1	GYS_DROME	Q9vfc8	drosophila
c 412	7	0.5	548	1	PPOC_TOBAC	Q24163	nicotiana t	c 485	7	0.5	692	1	PPK_ACIBA	Q9x4m8	acinetobact
c 413	7	0.5	552	1	ATPA_DROME	P35381	drosophila	c 486	7	0.5	704	1	MM09_CANFA	Q18733	canis famil
c 414	7	0.5	554	1	C971_PEA	Q43078	pisum sativ	c 487	7	0.5	704	1	SYNL_RAT	P09951	rattus norv
c 415	7	0.5	554	1	YJJK_ECOLI	P37797	escherichia	488	7	0.5	706	1	DREB_MOUSE	Q9qxs6	mus musculus
c 416	7	0.5	555	1	WEEL_XENLA	P47817	xenopus lae	489	7	0.5	707	1	DREB_RAT	Q07266	rattus norv
c 417	7	0.5	556	1	SYR_BACSU	P46906	bacillus su	490	7	0.5	708	1	GBF_DICDI	P36417	dictyostell
c 418	7	0.5	556	1	SYR_LISIN	Q927c2	listeria in	c 491	7	0.5	712	1	NRDD_ECOLI	P28903	escherichia
c 419	7	0.5	556	1	SYR_LISMO	Q8y493	listeria mo	492	7	0.5	716	1	DVL3_HUMAN	Q92997	homo sapien
c 420	7	0.5	556	1	YJJK_HAEIN	P45127	haemophilus	493	7	0.5	716	1	DVL3_MOUSE	Q61062	mus musculus
c 421	7	0.5	557	1	ACEA_YEAST	P28240	saccharomyc	494	7	0.5	717	1	MCCA_MOUSE	Q99mr8	mus musculus
c 422	7	0.5	557	1	FACC_RAT	Q35870	rattus norv	c 495	7	0.5	719	1	PBPA_STRPN	Q04707	streptococc
c 423	7	0.5	561	1	Y423_MYCPN	P75174	mycoplasma	496	7	0.5	732	1	TAU_MOUSE	P10637	mus musculus
c 424	7	0.5	562	1	4CL2_SOYBN	P31687	glycine max	497	7	0.5	735	1	CNG1_CHICK	Q90805	gallus gall
c 425	7	0.5	568	1	ZRF1_HUMAN	Q99543	homo sapien	498	7	0.5	740	1	POLB_MAIZE	P15718	zea mays (m
c 426	7	0.5	571	1	XYNC_PSEFL	P23031	pseudomonas	499	7	0.5	741	1	IF2_BACST	P04766	bacillus st
c 427	7	0.5	572	1	ERGI_MOUSE	P52019	mus musculus	500	7	0.5	741	1	TLE4_RAT	Q07141	rattus norv
c 428	7	0.5	573	1	ERGI_RAT	P52020	rattus norv	501	7	0.5	748	1	CAO_YEAST	P13711	saccharomyc
c 429	7	0.5	575	1	MIS_BOVIN	P03972	bos taurus	c 502	7	0.5	749	1	YU20_HUMAN	Q9y3t9	homo sapien
c 430	7	0.5	575	1	YOY6_CAEEL	P34664	caenorhabdi	c 503	7	0.5	750	1	PBPX_STRPN	P14677	streptococc
c 431	7	0.5	583	1	NH31_CAEEL	Q18192	caenorhabdi	504	7	0.5	758	1	L2DT_DROME	Q24371	drosophila
c 432	7	0.5	588	1	PR28_YEAST	P23394	saccharomyc	c 505	7	0.5	758	1	VKGC_RAT	Q88496	rattus norv
c 433	7	0.5	589	1	ARSE_HUMAN	P51690	homo sapien	c 506	7	0.5	761	1	SEPR_MOUSE	P97321	mus musculus
c 434	7	0.5	591	1	FACC_MOUSE	P50652	mus musculus	507	7	0.5	766	1	TLE4_HUMAN	Q04737	homo sapien
c 435	7	0.5	593	1	DCEL_MOUSE	P48318	mus musculus	508	7	0.5	766	1	TLE4_MOUSE	Q62441	mus musculus
c 436	7	0.5	593	1	DCEL_RAT	P18088	rattus norv	509	7	0.5	770	1	ACE2_YEAST	P21192	saccharomyc
c 437	7	0.5	595	1	YE72_HUMAN	Q9nx95	homo sapien	510	7	0.5	771	1	TLE3_MOUSE	Q08122	mus musculus
c 438	7	0.5	599	1	TYPA_HELPY	Q9z1z3	helicobacte	511	7	0.5	772	1	TLE3_HUMAN	Q04726	homo sapien
c 439	7	0.5	599	1	TYPA_HELPY	O25225	helicobacte	512	7	0.5	778	1	EXP7_STRPN	P35957	streptococc
c 440	7	0.5	602	1	CRK_DAUCA	P33681	daucus caro	c 513	7	0.5	782	1	BICD_DROME	P16568	drosophila
c 441	7	0.5	603	1	NU5M_GORGO	P03917	gorilla gor	c 514	7	0.5	782	1	Y044_UREPA	Q9pral	ureaplasma
c 442	7	0.5	603	1	VEL_HPV21	P50759	human papil	515	7	0.5	785	1	Y043_METJA	Q60348	methanococc
c 443	7	0.5	604	1	VEL_HPV19	Q02048	human papil	516	7	0.5	786	1	ST5B_HUMAN	P51692	homo sapien
c 444	7	0.5	604	1	VEL_HPV25	Q02049	human papil	517	7	0.5	787	1	ST5B_BOVIN	Q9tum3	bos taurus
c 445	7	0.5	605	1	VEL_HPV14	P36721	human papil	c 518	7	0.5	787	1	ST5B_PIG	Q9tuz0	sus scrofa
c 446	7	0.5	605	1	VEL_HPV20	P50758	human papil	c 519	7	0.5	791	1	K6PP_RABIT	P47859	oryctolagus
c 447	7	0.5	607	1	PPO_VITVI	P43311	vititis vinif	c 520	7	0.5	792	1	SYL_MYCGE	P47508	mycoplasma
c 448	7	0.5	616	1	VGFI_HUMAN	Q15240	homo sapien	521	7	0.5	794	1	ST5A_BOVIN	Q95115	bos taurus
c 449	7	0.5	617	1	RNC1_YEAST	P33753	saccharomyc	522	7	0.5	794	1	ST5A_HUMAN	P42229	homo sapien
c 450	7	0.5	619	1	HEX3_YEAST	P32828	saccharomyc	523	7	0.5	794	1	ST5A_SHEEP	P42231	ovis aries
c 451	7	0.5	619	1	SELB_HAEIN	P43927	haemophilus	c 524	7	0.5	797	1	S6A5_HUMAN	Q9y345	homo sapien
c 452	7	0.5	626	1	HCYB_EURCA	Q9nth9	eurytelma c	c 525	7	0.5	798	1	ITB0_XENLA	Q9y345	homo sapien
c 453	7	0.5	627	1	S6AB_MOUSE	P31650	mus musculus	c 526	7	0.5	798	1	ITB1_XENLA	P12606	xenopus lae
c 454	7	0.5	627	1	S6AB_RAT	P31647	rattus norv	527	7	0.5	799	1	ST5A_PIG	Q9tuz1	sus scrofa
c 455	7	0.5	628	1	RA21_SCHPO	P30776	schizosacch	c 528	7	0.5	803	1	ITB1_CHICK	P07228	gallus gall
c 456	7	0.5	629	1	DNAK_PAVLU	P30722	pavlova lut	529	7	0.5	807	1	PTNM_HUMAN	Q9y2r2	homo sapien

c 530	7	0.5	817	1	PMT1_YEAST	P33775 saccharomyc	603	7	0.5	1211	1	PKP4_HUMAN	Q99569 homo sapien
531	7	0.5	821	1	LIN5_CAEEL	P45970 caenorhabdi	604	7	0.5	1217	1	ITN1_RAT	Q9wve9 rattus norv
532	7	0.5	827	1	PLSB_MOUSE	G61586 mus musculu	605	7	0.5	1226	1	DSRA_HUMAN	P55265 homo sapien
533	7	0.5	828	1	PLSB_RAT	P97564 rattus norv	606	7	0.5	1227	1	B3A3_MOUSE	P16283 mus musculu
c 534	7	0.5	832	1	ANK3_HUMAN	P57078 homo sapien	607	7	0.5	1229	1	N121_HUMAN	Q9y2n3 homo sapien
535	7	0.5	835	1	FASD_ECOLI	P46000 escherichia	608	7	0.5	1233	1	NME3_HUMAN	Q14957 homo sapien
c 536	7	0.5	837	1	PSAC_YERPS	Q56983 yersinia ps	609	7	0.5	1237	1	NME3_RAT	Q00961 rattus norv
537	7	0.5	839	1	PHLD_BOVIN	P80109 bos taurus	610	7	0.5	1239	1	NME3_MOUSE	Q01098 mus musculu
c 538	7	0.5	845	1	KRUH_DROME	P08155 drosophila	c 611	7	0.5	1251	1	YQJ3_CAEEL	Q09550 caenorhabdi
539	7	0.5	855	1	ORP3_MOUSE	Q94825 mus musculu	c 612	7	0.5	1259	1	CAML_RAT	Q05895 rattus norv
c 540	7	0.5	860	1	BGL1_ASPAC	P48825 aspergillus	c 613	7	0.5	1260	1	CAML_MOUSE	P11627 mus musculu
c 541	7	0.5	868	1	TOP1_PASMO	Q9cn30 pasteurella	614	7	0.5	1276	1	PRDG_HUMAN	Q9haz2 homo sapien
542	7	0.5	872	1	SCD1_SCHPO	P09095 schizosacch	c 615	7	0.5	1280	1	MDR1_HUMAN	P08183 homo sapien
543	7	0.5	873	1	FAS2_DROME	P34082 drosophila	c 616	7	0.5	1360	1	GLI1_XENLA	Q91690 xenopus lae
544	7	0.5	891	1	VP4A_FOWPV	Q9j559 fowlpox vir	617	7	0.5	1435	1	LTEL_YEAST	P07866 saccharomyc
c 545	7	0.5	899	1	Y188_HUMAN	Q14693 homo sapien	618	7	0.5	1436	1	WIN1_SCHPO	Q74304 schizosacch
546	7	0.5	900	1	YB75_YEAST	P38321 saccharomyc	619	7	0.5	1482	1	NME2_MOUSE	Q01097 mus musculu
c 547	7	0.5	902	1	SYG1_YEAST	P40528 saccharomyc	620	7	0.5	1482	1	NME2_RAT	Q00960 rattus norv
c 548	7	0.5	912	1	ICA5_RABIT	Q28730 oryctolagus	621	7	0.5	1484	1	NME2_HUMAN	Q13224 homo sapien
c 549	7	0.5	917	1	ICA5_MOUSE	Q60625 mus musculu	c 622	7	0.5	1548	1	MDR_LEITA	P21441 leishmania
c 550	7	0.5	918	1	HXK1_MOUSE	P17710 mus musculu	c 623	7	0.5	1590	1	GCN2_YEAST	P15442 saccharomyc
c 551	7	0.5	918	1	HXK1_RAT	P05708 rattus norv	c 624	7	0.5	1592	1	YNY2_YEAST	P53855 saccharomyc
c 552	7	0.5	918	1	QALS_NEUCR	P11637 neurospora	c 625	7	0.5	1612	1	RRPO_PMMVJ	P89657 pepper mild
c 553	7	0.5	919	1	SY1_THEMA	P46213 thermotoga	c 626	7	0.5	1612	1	RRPO_PMMVJ	P29098 pepper mild
554	7	0.5	920	1	HELY_MYCLE	Q9zbd8 mycobacteri	c 627	7	0.5	1620	1	YO25_CAEEL	P34675 caenorhabdi
c 555	7	0.5	924	1	ICA5_HUMAN	Q9umf0 homo sapien	c 628	7	0.5	1658	1	YM67_YEAST	Q03661 saccharomyc
556	7	0.5	953	1	YMBG_YEAST	Q03516 saccharomyc	629	7	0.5	1666	1	CLH_SCHPO	Q10161 schizosacch
557	7	0.5	956	1	VP2_BTUV1	P05308 bluetongue	630	7	0.5	1678	1	CLH_DROME	P29742 drosophila
558	7	0.5	963	1	IR22_HUMAN	P48200 homo sapien	631	7	0.5	1704	1	VITI_FUNHE	Q90508 fundulus he
c 559	7	0.5	964	1	AMPN_RAT	P15684 rattus norv	632	7	0.5	1714	1	ITN1_MOUSE	Q920r4 mus musculu
c 560	7	0.5	965	1	AMPN_RABIT	P15541 oryctolagus	633	7	0.5	1804	1	YFA7_YEAST	P43583 saccharomyc
561	7	0.5	966	1	FIB1_PETMA	P02674 petromyzon	634	7	0.5	1812	1	BRC1_MOUSE	P48754 mus musculu
c 562	7	0.5	971	1	RECK_HUMAN	Q95980 homo sapien	c 635	7	0.5	1882	1	POL2_TPSVR	P25247 tomato ring
c 563	7	0.5	971	1	RECK_MOUSE	Q920j1 mus musculu	c 636	7	0.5	1886	1	POL2_COXIV	P19199 commelina y
c 564	7	0.5	974	1	CC15_YEAST	P27636 saccharomyc	637	7	0.5	1887	1	RPB1_DROME	P04052 drosophila
c 565	7	0.5	986	1	SM13_RAT	Q62839 rattus norv	638	7	0.5	1986	1	WA_EMENI	Q03149 emericella
c 566	7	0.5	993	1	GMI2_MOUSE	P40694 mus musculu	639	7	0.5	2012	1	DSCA_HUMAN	Q60469 homo sapien
c 567	7	0.5	1004	1	CARE_HUMAN	Q9b216 homo sapien	c 640	7	0.5	2035	1	EVPL_MOUSE	Q9d952 mus musculu
568	7	0.5	1005	1	GCP_CAEEL	Q10663 caenorhabdi	c 641	7	0.5	2109	1	PKS1_ASPPA	Q12053 aspergillus
569	7	0.5	1010	1	WNT5_DROME	P28466 drosophila	642	7	0.5	2161	1	SHK1_HUMAN	Q9y566 homo sapien
c 570	7	0.5	1013	1	YT93_DEIRA	Q9ry19 desinococcus	643	7	0.5	2167	1	SHK1_RAT	Q9wv48 rattus norv
571	7	0.5	1015	1	ITA4_DROME	Q9v7a4 drosophila	644	7	0.5	2181	1	STCA_EMENI	Q12397 emericella
c 572	7	0.5	1025	1	KQ09_HUMAN	Q9hoa0 homo sapien	c 645	7	0.5	2404	1	SON_MOUSE	Q9q47 mus musculu
573	7	0.5	1032	1	YQ53_CAEEL	Q09249 caenorhabdi	c 646	7	0.5	2410	1	POLI_BAYMJ	Q01206 barley yell
c 574	7	0.5	1033	1	ITAB_MOUSE	Q9qum0 mus musculu	c 647	7	0.5	2412	1	POLI_BAYMG	Q04574 barley yell
575	7	0.5	1050	1	EX5B_CHLPN	Q9z7g7 chlamydia p	648	7	0.5	2476	1	ATRX_MOUSE	Q61587 mus musculu
576	7	0.5	1050	1	ULX1_HUMAN	Q75385 homo sapien	649	7	0.5	2492	1	ATRX_HUMAN	P46100 homo sapien
c 577	7	0.5	1051	1	TFIA_MOUSE	Q64127 mus musculu	c 650	7	0.5	2555	1	FAFY_HUMAN	Q00507 h probable
578	7	0.5	1051	1	ULX1_MOUSE	Q70405 mus musculu	c 651	7	0.5	2594	1	7LES_DROVI	P20806 drosophila
579	7	0.5	1060	1	YNI8_YEAST	P53836 saccharomyc	c 652	7	0.5	2704	1	G168_PARPR	P17053 paramecium
580	7	0.5	1063	1	POLS_RUBVH	P21480 rubella vir	c 653	7	0.5	2715	1	G156_PARPR	P13837 paramecium
581	7	0.5	1063	1	POLS_RUBVT	P07566 rubella vir	654	7	0.5	2920	1	CLR2_MOUSE	Q9r0m0 mus musculu
c 582	7	0.5	1065	1	KDGI_HUMAN	Q75912 homo sapien	655	7	0.5	2923	1	CLR2_HUMAN	Q9hcu4 homo sapien
c 583	7	0.5	1075	1	RPOB_WAIZE	P16023 zea mays (m	656	7	0.5	3010	1	POLG_HCVTW	P29846 h genome po
c 584	7	0.5	1075	1	RPOB_ORISA	P12091 oryza sativ	657	7	0.5	3027	1	POLG_PYFV1	Q05057 parsnip yel
c 585	7	0.5	1076	1	RPOB_WHEAT	Q9xps7 triticum ae	658	7	0.5	3066	1	ATM_MOUSE	Q62388 mus musculu
c 586	7	0.5	1099	1	CARB_THEMA	Q9wz27 thermotoga	659	7	0.5	3178	1	YS89_CAEEL	Q09624 caenorhabdi
c 587	7	0.5	1121	1	ALP1_SCHPO	Q10197 schizosacch	c 660	7	0.5	3255	1	POLG_LMVO	P13199 l genome po
c 588	7	0.5	1121	1	DDX8_ARATH	Q38953 arabidopsis	c 661	7	0.5	3255	1	POLG_LMVE	P89876 l genome po
c 589	7	0.5	1121	1	EX5C_HAEIN	P44945 haemophilus	c 662	7	0.5	3313	1	CLR3_RAT	Q88278 rattus norv
590	7	0.5	1121	1	YE06_YEAST	P32644 saccharomyc	663	7	0.5	3680	1	DMD_CANFA	Q97592 canis famil
591	7	0.5	1127	1	Y855_TREPA	Q83827 treponema p	664	7	0.5	3703	1	ABF1_HUMAN	Q15911 homo sapien
c 592	7	0.5	1134	1	CYA1_BOVIN	P19754 bos taurus	c 665	7	0.5	3703	1	ABF1_MOUSE	Q15911 homo sapien
c 593	7	0.5	1138	1	DPS2_YEAST	P53037 saccharomyc	666	7	0.5	3726	1	ABF1_MOUSE	Q61329 mus musculu
c 594	7	0.5	1157	1	C8AA_BACUK	Q45704 bacillus th	c 667	7	0.5	3859	1	RPOA_LELV	Q04561 lelystad vi
c 595	7	0.5	1157	1	PBX1_PICPA	P46463 pichia past	668	7	0.5	3866	1	HRX_MOUSE	P55200 mus musculu
c 596	7	0.5	1163	1	ITAX_HUMAN	P20702 homo sapien	c 669	7	0.5	4036	1	RRPL_DUGBV	Q66431 dugbe virus
c 597	7	0.5	1165	1	TRA1_CAEBR	Q17308 caenorhabdi	670	7	0.5	4085	1	RRPA_CVH22	Q05002 human coron
598	7	0.5	1167	1	WCL_NEUCR	Q01371 neurospora	c 671	7	0.5	4367	1	DYHC_NEUCR	P45443 neurospora
c 599	7	0.5	1173	1	GIGA_ARATH	Q9sqd2 arabidopsis	c 672	7	0.5	4829	1	BIR6_HUMAN	Q9nr09 homo sapien
c 600	7	0.5	1173	1	GIGA_ORISA	Q9aw17 oryza sativ	c 673	7	0.5	5038	1	RYR1_HUMAN	P21817 homo sapien
601	7	0.5	1176	1	CARB_TRICU	P46056 trichosporo	c 674	7	0.5	6359	1	BACC_BACLI	Q68008 b bacitraci
602	7	0.5	1178	1	ITA2_MOUSE	Q62469 mus musculu	c 675	6	0.5	10	1	TEMK_RANTE	P56923 rana tempor

c 822	100	0.5	6	895	6	0.5	112	1	CYC_CRION	P00077 crithidia o
823	100	0.5	6	896	6	0.5	112	1	MYRL_MVRPI	Q07932 myrmecia pi
c 824	100	0.5	6	897	6	0.5	112	1	YE26_AQUAE	Q56349 paracoccus
c 825	101	0.5	6	898	6	0.5	112	1	CCB2_ECOLI	P02781 rattus norv
c 826	101	0.5	6	899	6	0.5	112	1	CDB6_ECOLI	Q9msm6 euphorbia e
827	101	0.5	6	900	6	0.5	112	1	HL16_ENSMI	P11301 bovine papi
828	101	0.5	6	901	6	0.5	113	1	MES1_SCHPO	P46966 homo sapien
829	101	0.5	6	902	6	0.5	113	1	CH10_STRCO	Q29036 sus scrofa
c 830	102	0.5	6	903	6	0.5	113	1	HSP2_HUMAN	Q4775 lactococcus
c 831	102	0.5	6	904	6	0.5	113	1	VAM5_MOUSE	Q47413 escherichia
c 832	102	0.5	6	905	6	0.5	114	1	VAM5_RAT	P81580 cancer pagu
c 833	102	0.5	6	906	6	0.5	114	1	YC4_TYLCA	P81201 stytonychia
c 834	102	0.5	6	907	6	0.5	114	1	YI82_ECOLI	P81196 stytonychia
835	103	0.5	6	908	6	0.5	114	1	HE2_HUMAN	Q9php0 campylobact
836	103	0.5	6	909	6	0.5	114	1	RNS_CEPSI	P01697 ocytolagus
837	103	0.5	6	910	6	0.5	114	1	YD29_HAEIN	Q03461 nicotiana t
838	103	0.5	6	911	6	0.5	114	1	YF98_ARCFU	P52475 human herpe
839	104	0.5	6	912	6	0.5	114	1	GAST_RAT	Q9j5b7 fowlpox vir
c 840	104	0.5	6	913	6	0.5	115	1	RK25_PEA	Q9p9b7 ureaplasma
841	104	0.5	6	914	6	0.5	115	1	RNS_SAITA	Q95708 hyllobates l
842	104	0.5	6	915	6	0.5	115	1	YJD2_YEAST	Q63362 rattus norv
843	105	0.5	6	916	6	0.5	115	1	ANT3_CHICK	P05871 human immun
c 844	105	0.5	6	917	6	0.5	115	1	CRTC_PIG	Q9w1f0 protopterus
c 845	105	0.5	6	918	6	0.5	115	1	MT1_TETPI	P22384 simian immu
c 846	105	0.5	6	919	6	0.5	115	1	RM34_YEAST	P76243 escherichia
847	105	0.5	6	920	6	0.5	116	1	RS1A_YEAST	P53542 clarias gar
848	105	0.5	6	921	6	0.5	116	1	RS1B_YEAST	Q9y9p3 ictalurus p
c 849	105	0.5	6	922	6	0.5	116	1	SUGE_CITFR	O23758 cicier ariet
c 850	105	0.5	6	923	6	0.5	116	1	WUNI_SOLTU	P20686 oncorhynch
c 851	105	0.5	6	924	6	0.5	116	1	Y450_AQUAE	Q35262 oncorhynch
c 852	105	0.5	6	925	6	0.5	116	1	YRKD_BACSU	P20687 oncorhynch
c 853	106	0.5	6	926	6	0.5	116	1	KACB_RABIT	Q37108 oncorhynch
c 854	106	0.5	6	927	6	0.5	116	1	SECE_HAEIN	P11629 oncorhynch
c 855	106	0.5	6	928	6	0.5	116	1	THIO_PASMU	P20688 oncorhynch
c 856	106	0.5	6	929	6	0.5	116	1	VE7_HPV45	P25077 oncorhynch
857	106	0.5	6	930	6	0.5	116	1	YNT0_RHOCA	Q03252 salmo trutt
c 858	107	0.5	6	931	6	0.5	116	1	BOMB_BOMVA	P19548 human immun
c 859	107	0.5	6	932	6	0.5	116	1	HMB4_TRIGR	P35960 human immun
860	107	0.5	6	933	6	0.5	116	1	HXAD_AMBME	Q8r980 fusobacteri
c 861	107	0.5	6	934	6	0.5	116	1	PRVB_ESOLU	P36340 simian immu
c 862	107	0.5	6	935	6	0.5	116	1	PL17_YEAST	Q9x1k3 arabidopsis
c 863	107	0.5	6	936	6	0.5	116	1	RLA2_TRYBB	O50717 haemophilus
c 864	107	0.5	6	937	6	0.5	116	1	THIO_PORPU	P53222 saccharomyc
c 865	107	0.5	6	938	6	0.5	117	1	THIO_PORYE	P03981 cainan croc
c 866	108	0.5	6	939	6	0.5	117	1	KREJ_CHICK	P01785 canis famil
c 867	108	0.5	6	940	6	0.5	117	1	KREJ_COTJA	P19180 carassius a
868	108	0.5	6	941	6	0.5	117	1	KV50_MOUSE	P01749 mus musculu
869	108	0.5	6	942	6	0.5	117	1	KV55_MOUSE	P01750 mus musculu
870	108	0.5	6	943	6	0.5	117	1	KV5T_MOUSE	O29138 archaeoglob
c 871	108	0.5	6	944	6	0.5	117	1	THIO_BUCAP	P22850 chromatium
c 872	108	0.5	6	945	6	0.5	117	1	VIB3_AGR5	P56841 mus musculu
c 873	108	0.5	6	946	6	0.5	117	1	YC1H_ECOLI	P01212 cyprinus ca
c 874	108	0.5	6	947	6	0.5	118	1	YD08_METJA	P30983 ctenopharyn
c 875	109	0.5	6	948	6	0.5	118	1	COXE_HUMAN	P37037 hypophthalm
c 876	109	0.5	6	949	6	0.5	118	1	COXE_RABIT	P10117 laticauda c
c 877	109	0.5	6	950	6	0.5	118	1	NIEW_AZOCA	P77986 staphylococ
c 878	109	0.5	6	951	6	0.5	118	1	SZ13_MOUSE	Q00190 escherichia
c 879	109	0.5	6	952	6	0.5	118	1	YG1C_YEAST	P39111 saccharomyc
c 880	109	0.5	6	953	6	0.5	118	1	YJGZ_ECOLI	P01887 mus musculu
c 881	110	0.5	6	954	6	0.5	119	1	COX4_BACSU	Q9chq3 lactococcus
c 882	110	0.5	6	955	6	0.5	119	1	ETS2_LYVIA	P76248 escherichia
c 883	110	0.5	6	956	6	0.5	119	1	FLIN_CAUCR	P76259 escherichia
c 884	110	0.5	6	957	6	0.5	119	1	KV15_RABIT	Q02009 lactococcus
c 885	110	0.5	6	958	6	0.5	119	1	RL24_THETH	Q9muq9 mesostigma
886	110	0.5	6	959	6	0.5	120	1	Y063_RICPR	Q8tyc7 methanopyru
887	110	0.5	6	960	6	0.5	120	1	YIDQ_ECOLI	P06926 bovine papi
c 888	111	0.5	6	961	6	0.5	120	1	COXE_MOUSE	P15267 homo sapien
c 889	111	0.5	6	962	6	0.5	121	1	COXE_RAT	Q99978 porphyra pu
890	111	0.5	6	963	6	0.5	121	1	FLIE_BORBU	P45648 coxiella bu
891	111	0.5	6	964	6	0.5	121	1	N3M_ASCSU	Q00144 ictaluriid h
892	111	0.5	6	965	6	0.5	121	1	YH07_YEAST	P76334 escherichia
893	111	0.5	6	966	6	0.5	121	1	YR2C_BACSU	P80090 lymnaea sta
894	112	0.5	6	967	6	0.5	122	1	CFP6_MYCTU	

968 6 0.5 122 1 YRHF_BACSU
 c 969 6 0.5 123 1 C59A_MOUSE
 c 970 6 0.5 123 1 CRCB_PYRHO
 971 6 0.5 123 1 ELI2_PHYCR
 972 6 0.5 123 1 ELI3_PHYCR
 c 973 6 0.5 123 1 PFDB_AERPE
 c 974 6 0.5 123 1 RBFA_RALSO
 975 6 0.5 123 1 RBFA_SPIOL
 976 6 0.5 123 1 RBS1_NOCSE
 977 6 0.5 124 1 HVL1D_HUMAN
 c 978 6 0.5 124 1 MAG5_HUMAN
 c 979 6 0.5 124 1 RS16_RHIME
 c 980 6 0.5 124 1 URE2_BACSU
 981 6 0.5 124 1 VATE_NEUCR
 c 982 6 0.5 124 1 VE4_HPV57
 983 6 0.5 125 1 HV1F_HUMAN
 c 984 6 0.5 125 1 RS13_RICPR
 c 985 6 0.5 126 1 CCB3_ECOLI
 c 986 6 0.5 126 1 CCB4_ECOLI
 c 987 6 0.5 126 1 CLP2_DROME
 c 988 6 0.5 126 1 PRO1_DICDI
 c 989 6 0.5 126 1 RBS1_ACECL
 990 6 0.5 126 1 SECE_VIBCH
 991 6 0.5 126 1 VXS1_BPAPS
 c 992 6 0.5 127 1 CY3A_DESAF
 993 6 0.5 127 1 GTH1_ANGJA
 c 994 6 0.5 127 1 LY6D_MOUSE
 c 995 6 0.5 127 1 RL17_XANCP
 c 996 6 0.5 127 1 YLH1_ECOLI
 c 997 6 0.5 128 1 CION_CIOIN
 998 6 0.5 128 1 RNP_MAIZE
 999 6 0.5 128 1 RNP_PREEN
 1000 6 0.5 128 1 YBM6_YEAST

ALIGNMENTS

RESULT 1
 MEA6_HUMAN STANDARD; PRT; 804 AA.
 ID MEA6_HUMAN STANDARD; PRT; 804 AA.
 AC O15320: O00169;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Meningioma-expressed antigen 6/11 (MEA6) (MEAL1).
 GN MEA6 OR MGEAL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS MEA6 AND MEAL1).
 RC TISSUE=Meningioma;
 RX MEDLINE=97472454; PubMed=9356211;
 RA Heckel D., Brass N., Fischer U., Blin N., Steudel I., Tueresi O.,
 RA Fackler O., Zang K.D., Meese E.;
 RT "cDNA cloning and chromosomal mapping of a predicted coiled-coil
 RT proline-rich protein immunogenic in meningioma patients.";
 RL Hum. Mol. Genet. 6:2031-2041(1997).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEA6 (SHOWN HERE) AND MEAL1; MAY
 CC BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED IN BRAIN, MUSCLE
 CC AND CRANIAL SKIN.
 CC -1- DISEASE: AUTOANTIGEN FOUND IN BENIGN MENINGIOMA.
 CC -1- CAUTION: AT LEAST 9 PSEUDOGENES HAVE BEEN FOUND ON DIFFERENT
 CC CHROMOSOMES, INCLUDING CHROMOSOMES 2, 3, 6, 7, 9, 10, 12, 13 AND
 CC 18.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X07782; CAA30615.1; -
 CC PIR: S01852; HVRKCL.
 CC HSSP: P01842; 7FAB.
 CC InterPro: IPR003006; Iq_MHC.
 CC InterPro: IPR003597; Iq-cl.
 CC InterPro: IPR003600; Iq-like.
 CC Pfam: PF00047; Iq; 3.
 CC SMART: SM00410; Iq-like; 1.

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U94780; AAB86593.1; -
 CC EMBL: U73682; AAB86589.1; ALT_INIT.
 CC DR Genew: HGNC:7057; MGEA6.
 CC DR MIM: 602132; -
 CC KW Antigen; Coiled coil; Alternative splicing.
 FT DOMAIN 87 273 COILED COIL (POTENTIAL).
 FT DOMAIN 321 501 COILED COIL (POTENTIAL).
 FT DOMAIN 553 558 POLY-GLY.
 FT DOMAIN 502 804 PRO-RICH.
 FT VARSPIC 514 556 MISSING (IN ISOFORM MEAL1).
 FT CONFLICT 198 198 P -> O (IN REF. 1; AAB86589).
 FT CONFLICT 360 360 E -> O (IN REF. 1; AAB86589).
 FT CONFLICT 589 589 L -> P (IN REF. 1; AAB86589).
 FT CONFLICT 594 594 F -> S (IN REF. 1; AAB86589).
 FT CONFLICT 699 699 V -> I (IN REF. 1; AAB86589).
 FT CONFLICT 738 738 G -> R (IN REF. 1; AAB86589).
 SQ SEQUENCE 804 AA; 90943 MW; 3F088799E0883EDE CRC64;
 Alignment Scores:
 Pred. No.: 0.293 Length: 804
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.76% Indels: 0
 DB: 1 Gaps: 0
 US-09-818-990b-1 (1-3963) x MEA6_HUMAN (1-804)
 QY 161 GCGGAGGCCAAGATGACCTCCAGATCTTT 190
 |||||
 DB 188 AlaGluAlaLysMetThrPheGlnIlePhe 197
 |||||
 RESULT 2
 HVC3_HETFR
 ID HVC3_HETFR STANDARD; PRT; 393 AA.
 AC P23086;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig heavy chain C region (Clone 6121) (Fragment).
 OS Heterodontus francisci (Horn shark).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
 OC Heterodontidae; Heterodontus.
 OX NCBI_TaxID=7792;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RX MEDLINE=68328985; PubMed=3138109;
 RA Kokubu F., Hinds K., Litman R., Shambloot M.J., Litman G.W.;
 RT "Complete structure and organization of immunoglobulin heavy chain
 RT constant region genes in a phylogenetically primitive vertebrate.";
 RL EMBO J. 7:1979-1988(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X07782; CAA30615.1; -
 CC PIR: S01852; HVRKCL.
 CC HSSP: P01842; 7FAB.
 CC InterPro: IPR003006; Iq_MHC.
 CC InterPro: IPR003597; Iq-cl.
 CC InterPro: IPR003600; Iq-like.
 CC Pfam: PF00047; Iq; 3.
 CC SMART: SM00410; Iq-like; 1.

```
DR SMART; SM00407; IGcl; 2.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 393 AA; 43081 MW; 4E44D076972F18B5 CRC64;

Alignment Scores:
Pred. No.: 3 37 Length: 393
Score: 9.00 Matches: 9
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.68%
Indels: 0
DB: 1

US-09-818-990B-1 (1-393) x HVC3_HETFR (1-393)

QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706
| | | | | | | | | | | | | | | | | |
Db 251 LeuProThrProLeuArgAlaSerIle 259

RESULT 3
HVC2_HETFR
ID HVC2_HETFR STANDARD; PRT; 438 AA.
AC P23085;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region (Clone 12022) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Littman R., Shambloott M.J., Littman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07784; CAA30617.1; -.
CC PIR; S00980; HVRKC2.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR003006; Iq_MHC.
CC InterPro; IPR003597; Iq_cl.
CC InterPro; IPR003600; Iq_like.
CC Pfam; PF00047; Iq; 4.
CC SMART; SM00410; IG_like; 2.
CC SMART; SM00407; IGcl; 2.
CC PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 438 AA; 47904 MW; 6E67A9C5BBF7AA7B CRC64;

Alignment Scores:
Pred. No.: 3 35 Length: 438
Score: 9.00 Matches: 9
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.68%
Indels: 0
DB: 1

US-09-818-990B-1 (1-3963) x HVC2_HETFR (1-438)

QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706
| | | | | | | | | | | | | | | | | |
Db 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 4
HVC3_HETFR
ID HVC3_HETFR STANDARD; PRT; 438 AA.
AC P23087;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, secreted form (Clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Littman R., Shambloott M.J., Littman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07781; CAA30613.1; ALT_INIT.
CC PIR; S01853; HVRKCS.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR003006; Iq_MHC.
CC InterPro; IPR003597; Iq_cl.
CC InterPro; IPR003600; Iq_like.
CC Pfam; PF00047; Iq; 4.
CC SMART; SM00410; IG_like; 2.
CC SMART; SM00407; IGcl; 2.
CC PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 1 99 CH1.
FT DOMAIN 100 205 CH2.
FT DOMAIN 206 308 CH3.
FT DOMAIN 309 418 CH4.
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 438 AA; 47920 MW; 9BDF37C3BBD7D15B CRC64;

Alignment Scores:
Pred. No.: 3.35 Length: 438
Score: 9.00 Matches: 9
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.68%
Indels: 0
Gaps: 0
DB: 1

US-09-818-990B-1 (1-3963) x HVCS_HETER (1-438)

QY 1680 CTCGCCACCCACAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerille 304

RESULT 5

HVCM_HETER
ID HVCM_HETER STANDARD; PRT; 461 AA.
AC P23088:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, membrane-bound form (Clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Spleen;
RA MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shambloft M.J., Litman G.W.;
RT *Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.*;
RL EMBO J. 7:1979-1988(1988).
RN [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X07781; CAA30614.1; ALT_SEQ.
DR PIR; S01854; HVRKCO.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003597; Ig_cl.
DR InterPro; IPR003600; Ig_like.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00410; Ig_like; 2.
DR SMART; SM00407; IGC1; 2.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane.

FT NON_TER 1 1
FT DOMAIN 1 99 CH1.
FT DOMAIN 100 205 CH2.
FT DOMAIN 206 308 CH3.
FT DOMAIN 309 418 CH4.
FT TRANSHEM 438 458 POTENTIAL.
FT CARBOHYD 164 164
FT CARBOHYD 200 200
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 461 AA; 50762 MW; 029E91C8D5DF911 CRC64;

Alignment Scores:
Pred. No.: 3.34 Length: 461
Score: 9.00 Matches: 9
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.68%
Indels: 0
Gaps: 0
DB: 1

US-09-818-990B-1 (1-3963) x HVCM_HETER (1-461)

QY 1680 CTCGCCACCCACAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerille 304

RESULT 6

UL33_HSV11
ID UL33_HSV11 STANDARD; PRT; 130 AA.
AC P10217:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Protein UL33.
GN UL33.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1";
RL J. Gen. Virol. 69:1531-1574(1988).
RN [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D10879; BAA01679.1; -
DR EMBL; X14112; CAA32308.1; -
DR EMBL; M62932; AAA45829.1; -
DR PIR; F30085; WMBEH3.
DR InterPro; IPR005208; Herpes_UL33.
DR Pfam; PF03581; Herpes_UL33; 1.
FT MUTAGEN 17 17 I->D: TEMPERATURE SENSITIVE.
SQ SEQUENCE 130 AA; 14437 MW; D04CFDA7A3C585D3 CRC64;

Alignment Scores:
Pred. No.: 39.7 Length: 130
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.63%
Indels: 0
Gaps: 0
DB: 1

US-09-818-990B-1 (1-3963) x UL33_HSV11 (1-130)

QY 2519 ATGCGAGGAGAGAGGAGCAACA 2496

Db 1 MetAlaGlyArgGluGlyArgThr 8

RESULT 7

Y006_DEIRA .STANDARD; PRT; 227 AA.

AC Q9VE3;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical metal-dependent hydrolase DR0006.

GN DR0006.

OS Deinococcus radiodurans.

OC Bacteria; thermus/Deinococcus group; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OX NCBI_TaxID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RL;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,

RA Kachrova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans RL."

RL Science 286:1571-1577(1999).

CC -1- SIMILARITY: BELONGS TO THE UPF0173 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; A5001864; AAF09599.1; ALT_INIT.

DR TIGR; DR0006;

KW Hypothetical protein; Hydrolase; Complete proteome.

SQ SEQUENCE 227 AA; 24488 MW; EF3B786678940265 CRC64;

Alignment Scores:

Pred. No.:	38.4	Length:	227
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.61%	Indels:	0
DB:	1	Gaps:	0

US-09-818-990B-1 (1-3963) x Y006_DEIRA (1-227)

QY 3519 GATCCTGGAGAACTACAGAACTG 3542

Db 220 AspProGlyGluThrGluLeu 227

RESULT 8

ALSR_BACSU

ID ALSR_BACSU .STANDARD; PRT; 302 AA.

AC Q04778;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ALS operon regulatory protein.

GN ALSR.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93286000; PubMed=7695336;
RA Renna M.C., Najmudin N., Winik L.R., Zahler S.A.;
RT "Regulation of the Bacillus subtilis alss, alssD, and alssR genes
RT involved in post-exponential-phase production of acetoin.";
RL J. Bacteriol. 175:3863-3875(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98015417; PubMed=9353933;
RA Presecan E., Moszer I., Boursier L., Cruz Ramos H.C., De La Fuente V.,
RA Hulio M.-F., Lelong C., Schleich S., Skowaska A., Song B.H.,
RA Villani G., Kunst F., Danchin A., Glaser P.;
RT "The Bacillus subtilis genome from gerBC (311 degrees) to lscr (334
RT degrees).";
RL Microbiology 143:3313-3328(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: REGULATES THE EXPRESSION OF THE ALSSD OPERON FOR
CC ACETON BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; L04470; AAA2221.1;
EMBL; Z93767; CAB07788.1;
EMBL; Z99122; CAB15619.1;
PIR; A47126; A47126.
Subtilist; BG10470; alssR.
InterPro; IPR000847; HTH_LysR.

DR InterPro: IPR005119; LysR_subst.
 DR Pfam: PF00126; HTH_1; 1.
 DR Pfam: PF03466; LysR_substrate; 1.
 DR PRINTS: PR00039; HTHLYSR.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
 KW Transcription regulation; DNA-binding; Acetoin biosynthesis;
 KW Complete proteome.
 FT DNA_BIND 18 37 H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 302 AA; 34333 MW; BCA4DF5CF95A3E33 CRC64;

Alignment Scores:
 Pred. No.: 37.7 Length: 302
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x ALSR_BACSU (1-302)
 QY 659 CTGCTATCCGAGGATAGGATG 636
 |||||
 Db 232 LeuValSerAlaGlyIleGlyMet 239

RESULT 9
 COAA_MYCTU
 ID COAA_MYCTU STANDARD; PRT; 312 AA.
 AC O53440;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pantothenate kinase (EC 2.7.1.33) (Pantothenic acid kinase).
 GN COAA OR RV1092C OR MT1124 OR MTV017.45C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weldman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
 CC phosphopantothenate.
 CC -1- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -1- SIMILARITY: BELONGS TO THE PANTOTHENATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL021897; CAAL7208.1; -
 DR EMBL: AE006992; AAK45382.1; -
 DR TIGR: MT1124; -
 DR TuberculList; RV1092c; -
 DR InterPro: IPR001324; PRK.
 DR InterPro: IPR004566; Pank_bact.
 DR Pfam: PF00485; PRK; 1.
 DR TIGRFAMS: TIGR00554; pank_bact; 1.
 KW Transferase; Kinase; ATP-binding; Coenzyme A biosynthesis;
 KW Complete proteome.
 FT NP_BIND 97 104 ATP (POTENTIAL).
 SQ SEQUENCE 312 AA; 35656 MW; F075B543AE75788D CRC64;

Alignment Scores:
 Pred. No.: 37.6 Length: 312
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x COAA_MYCTU (1-312)
 QY 383 GAGCTGGTGGACTTCGAGGGTTA 360
 |||||
 Db 33 GluLeuValGlyLeuArgGlyLeu 40

RESULT 10
 HXDB_MOUSE
 ID HXDB_MOUSE STANDARD; PRT; 323 AA.
 AC P23813;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-d11 (Hox-4.6) (Hox-5.5).
 GN HOXD11 OR HOXD-11 OR HOX-4.6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91293104; PubMed=1676674;
 RA Izpisua-Belmonte J.-C., Falkenstein H., Dolle P., Renucci A.,
 RA Duboule D.;
 RT "Murine genes related to the Drosophila AbdB homeotic genes are
 RT sequentially expressed during development of the posterior part of
 RT the body."
 RL EMBO J. 10:2279-2289(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=94074553; PubMed=7902810;
 RA Gerard M., Duboule D., Zakany J.;
 RT "Structure and activity of regulatory elements involved in the
 RT activation of the Hoxd-11 gene during late gastrulation."
 RL EMBO J. 12:3539-3550(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV; TISSUE=Liver;
 RA Takada S., Cook M., Kramlauf R., McMahon A.P.;
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING DEVELOPMENT OF THE
 CC POSTERIOR PART OF THE BODY.
 CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.


```

-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X60761; CAA43173.1; -
DR EMBL; X60762; CAA43173.1; JOINED.
DR EMBL; X71422; CAA50553.1; -
DR EMBL; X60395; CAA42943.1; ALT_INIT.
DR PIR; S16318; S16318.
DR PIR; S40403; S40403.
DR HSP; PL4653; 1B72.
DR TRANSFAC; T01760; -
DR MGD; MGI:96203; Hoxd11.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 64 87 POLY-GLY.
FT DOMAIN 93 105 POLY-ALA.
FT DNA_BIND 251 310 HOMEBOX.
SQ SEQUENCE 323 AA; 33545 MW; B003D1347F5FE771 CRC64;

Alignment Scores:
Pred. No.: 37.5 Length: 323
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x HXDB_MOUSE (1-323)
QY 145 GGGCGCGTGAAGGAGCGGAGGC 168
|||||
DB 221 GlyAlaAlaGluGlyGlyGly 228

RESULT 11
NODI_RHISN
ID NODI_RHISN STANDARD; PRT; 343 AA.
AC P55476;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NODI OR Y4HF.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Frelberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE000076; AAB91694.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF0003439; ABC_transportr.
DR ProDom; PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFS; TIGR01288; nodI; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Nodulation; Membrane; ATP-binding; Transport; Plasmid.
FT NP_BIND 77 84 ATP (POTENTIAL).
SQ SEQUENCE 343 AA; 37917 MW; F49A7EC56E099A33 CRC64;

Alignment Scores:
Pred. No.: 37.4 Length: 343
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x NODI_RHISN (1-343)
QY 169 GGCCTCCGCTCTTCAGCGGCC 146
|||||
DB 319 GlyLeuArgLeuLeuGlnArgPro 326

RESULT 12
ODBB_RAT
ID ODBB_RAT STANDARD; PRT; 369 AA.
AC P35738;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 2-oxoisovalerate dehydrogenase beta subunit, mitochondrial precursor
DE (EC 1.2.4.4) (Branched-chain alpha-keto acid dehydrogenase component
DE beta chain (E1)) (BCKDH E1-beta) (Fragment).
GN BCKDHB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-54.
RX MEDLINE=93003326; PubMed=1390893;
RA Zhao Y., Kuntz M.J., Harris R.A., Crabb D.W.;
RT "Molecular cloning of the E1 beta subunit of the rat branched chain
RT alpha-ketoacid dehydrogenase.";
RL Biochim. Biophys. Acta 1132:207-210(1992).
RN [2]
RP ERRATUM
RX MEDLINE=95143286; PubMed=7841205;
RA Zhao Y., Kuntz M.J., Harris R.A., Crabb D.W.;
RL Biochim. Biophys. Acta 1260:243-243(1995).
CC -!- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:
CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).
CC -!- CATALYTIC ACTIVITY: 3-methyl-2-oxobutanoate + lipoamide = S-(2-
CC methylpropanoyl)dihydroliipoamide + CO(2).
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M94040; AAT73899.1; -
 CC PIR; S28950; S28950.
 CC HSP; P09061; IQSO.
 CC InterPro; IPR000360; Transketolase.
 CC Pfam; PF02779; transket_pyr; 1.
 CC Pfam; PF02780; transketolase_C; 1.
 CC Oxidoreductase; Mitochondrion; Transit peptide.
 CC NON_TER 1
 CC TRANSIT <1 27 MITOCHONDRION.
 CC CHAIN 28 369 2-OXOISOVALERATE DEHYDROGENASE BETA
 CC SUBUNIT
 CC SEQUENCE 369 AA; 40561 MW; 33E7B84EAB8178FD CRC64;

Alignment Scores:
 Pred. No.: 37.2 Length: 369
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x ODBB_RAT (1-369)

QY 2289 GAGGCTTCCTTGCTCACTGCTCT 2266

DB 178 GLUALAPheAlaHisCysPro 185

RESULT 13

ID PEL2_ERWCA STANDARD; PRT; 374 AA.

AC P11431; Q06112; Q47469;

DT 01-OCT-1989 (Rel. 12, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Pectate lyase II precursor (EC 4.2.2.2) (PEL II) (PLB).

GN PEL2 OR PELB.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

OX NCBI_TaxID=554;

ON [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ER / IAM1068;

RX MEDLINE=93113068; PubMed=1369060;

RA Yoshida A., Matsuo Y., Kamio Y., Izaki K.;

RT "Molecular cloning and sequencing of the extracellular pectate lyase

II gene from Erwinia carotovora Er";

RL Biosci. Biotechnol. Biochem. 56:1596-1600(1992).

RN [2]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=Sp. atroseptica EC;

RX MEDLINE=87308030; PubMed=3040692;

RA Lei S.-P., Lin H.-C., Wang S.-S., Callaway J., Wilcox G.;

RT "Characterization of the Erwinia carotovora pelB gene and its product

pectate lyase";

RL J. Bacteriol. 169:4379-4383(1987).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Sp. atroseptica C18.

RX MEDLINE=95291436; PubMed=7773390;

RA Bartling S., Wegener C., Olsen O.;

RT "Synergism between Erwinia pectate lyase isoenzymes that depolymerize

both pectate and pectin";

RL Microbiology 141:873-881(1995).

CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.

CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give

oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at

their non-reducing ends.

CC

CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1. PLADES
 CC SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; S51475; AAC60422.1; -

CC EMBL; M17364; AAA24848.1; -

CC EMBL; X81847; CAA57440.1; -

CC HSP; P11073; IAIR.

CC InterPro; IPR002022; Amb_allergen.

CC Pfam; PF00544; pec_lyase; 1.

CC KW Lyase; Multigene family; Signal.

FT SIGNAL 1 22

FT CHAIN 23 374 PECTATE LYASE II.

FT DISULFID 93 176 BY SIMILARITY.

FT DISULFID 350 373 BY SIMILARITY.

FT ACT_SITE 239 239 POTENTIAL.

FT VARIANT 35 35 E -> D (IN STRAIN EC).

FT VARIANT 47 48 MK -> LQ (IN STRAIN EC).

FT VARIANT 58 58 Q -> K (IN STRAIN EC).

FT VARIANT 78 78 S -> N (IN STRAIN EC).

FT VARIANT 136 136 V -> L (IN STRAIN C18).

FT VARIANT 139 139 M -> I (IN STRAIN C18).

FT VARIANT 144 144 M -> I (IN STRAIN C18).

FT VARIANT 150 150 D -> H (IN STRAIN C18).

FT VARIANT 156 156 I -> V (IN STRAIN EC).

FT VARIANT 168 168 E -> K (IN STRAIN EC).

FT VARIANT 172 173 KN -> QS (IN STRAIN EC).

FT VARIANT 234 234 R -> S (IN STRAIN C18).

FT VARIANT 252 252 N -> T (IN STRAIN EC).

FT VARIANT 280 280 N -> I (IN STRAIN EC).

FT VARIANT 306 311 SPSDFA -> KPADES (IN STRAINS C18 AND EC).

FT VARIANT 314 314 K -> N (IN STRAIN EC).

FT VARIANT 319 319 K -> R (IN STRAIN EC).

FT VARIANT 322 322 S -> T (IN STRAIN EC).

FT VARIANT 325 325 I -> V (IN STRAIN EC).

FT VARIANT 329 329 N -> D (IN STRAINS C18 AND EC).

FT VARIANT 338 339 SI -> AV (IN STRAIN C18).

FT SEQUENCE 374 AA; 40380 MW; FAE3FD00ACD29ED5 CRC64;

Alignment Scores:

Pred. No.: 37.2 Length: 374
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x PEL2_ERWCA (1-374)

QY 3036 ATCACTGGTAGTCGATCAATGTC 3013

DB 258 IleThrolySerglyPheAsnVal 265

RESULT 14

SLGL_YEAST

ID SLGL_YEAST STANDARD; PRT; 378 AA.

AC P54867;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SLG1 protein precursor.

GN SLG1 OR YOR008C OR UNP378.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bettignies G., Bergez-Aulio P., Barthe C., Louvet O.,
RA Peyququet M.F., Morel C., Doignon F., Crouzet M.;
RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051599; PubMed=8896276;
RA Sterky F., Holmberg A., Petterson B., Uhlen M.;
RT "The sequence of a 30 kb fragment on the left arm of chromosome XV
RT from Saccharomyces cerevisiae reveals 15 open reading frames, five of
RT which correspond to previously identified genes.";
RL Yeast 12:1091-1095(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39481; AAC05862.1; -;
DR EMBL: U43491; AAC49488.1; -;
DR EMBL: Z74916; CAA99196.1; -;
DR SGD: S0005534; SLG1.
DR InterPro: IPR002889; WSC.
DR Pfam: PF01822; WSC; 1.
DR SMART: SM00321; WSC; 1.
DR Glycoprotein: Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 378 SLG1 PROTEIN.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 378 AA; 39270 MW; EEE164F2374CCCE3 CRC64;

Alignment Scores:
Pred. No.: 37.1 Length: 378
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x SLG1_YEAST (1-378)

QY 2795 ATTCATCATCTGATTCATCAACA 2772
|||||
DB 118 ILeSerSerSerAspSerSerThr 125

RESULT 15
SGAA_METEX STANDARD; PRT; 379 AA.
AC P55819;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Serine--glyoxylate aminotransferase (EC 2.6.1.45) (SGAT).
GN SGAA.
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Methylobacterium group; Methylobacterium.
OX NCBI_TaxID=408;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AM1 / NCIMB 9133;
RA MEDLINE=94193542; PubMed=8144463;
RA Chistoserdova L.V., Lidstrom M.E.;
RT "Genetics of the serine cycle in Methylobacterium extorquens AM1:
RT identification of sgaA and mtdA and sequences of sgaA, hprA, and
RT mtdA.";
RL J. Bacteriol. 176:1957-1968(1994).

CC -!- CATALYTIC ACTIVITY: L-serine + glyoxylate -> 3-hydroxypyruvate +
CC glycine.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: SERINE PATHWAY.
CC -!- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27235; -; NOT_ANNOTATED_CDS.
DR InterPro: IPR000192; Aminotransf.
DR Pfam: PF00266; aminotran_5; 1.
DR PROSITE: PS00595; AA_TRANSFER_CLASS_5; 1.
DR Transferase: Aminotransferase; Pyridoxal phosphate.
FT BINDING 178 178 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 379 AA; 40709 MW; BEE45F4DC932A3EF CRC64;

Alignment Scores:
Pred. No.: 37.1 Length: 379
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x SGAA_METEX (1-379)

QY 2832 AGCAATACACTTGCCTGGGGAT 2809
|||||
DB 73 SerAsnThrLeuAlaArgGlyAsp 80

Search completed: November 30, 2002, 18:44:38
Job time : 57 secs
```


GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:36:09 ; Search time 98.5 Seconds

(without alignments)

16579.994 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacagacagcataga.....tggagagtgtgaactttaa 3963

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09818990/runat_26112002_093222_22047/app_query.fasta_1.4103
-DB=SPREMBL_21 -OFMT=fastan -SUFFIX=oligo.rspt -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.140 @runat_26112002_093222_22047 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	770	58.5	1320 4	Q96kf5 homo sapien

2	494	37.5	507	4	Q96K90
3	19	1.4	385	4	Q9UOF5
4	19	1.4	404	4	Q9Y3E9
5	19	1.4	772	4	Q9Y2J6
6	19	1.4	1106	4	Q8WX93
7	16	1.2	379	11	Q9CWM1
8	16	1.2	492	11	Q9ET54
9	9	0.7	156	17	Q9Y6G5
10	9	0.7	239	17	Q8TXM0
11	9	0.7	243	2	Q9EV74
12	9	0.7	411	3	Q9C196
13	9	0.7	490	16	Q9K742
14	9	0.7	587	11	Q88831
15	9	0.7	663	10	Q9LT87
16	9	0.7	897	3	Q08886
17	9	0.7	1756	3	Q12193
18	9	0.7	1884	5	Q9NHW2
19	9	0.7	3469	5	Q9U412
20	9	0.7	3502	5	Q9VYJ9
21	8	0.6	48	5	Q9VWS6
22	8	0.6	102	16	Q911U1
23	8	0.6	103	12	Q99FW9
24	8	0.6	107	16	Q9RC48
25	8	0.6	110	13	O42426
26	8	0.6	120	8	O99132
27	8	0.6	120	17	O82U19
28	8	0.6	122	16	O9KTN9
29	8	0.6	122	17	O58355
30	8	0.6	123	17	O8TY74
31	8	0.6	133	10	Q9M7N3
32	8	0.6	141	11	Q61692
33	8	0.6	142	10	O9M7N5
34	8	0.6	149	2	O51755
35	8	0.6	149	11	Q9D4M1
36	8	0.6	153	11	Q9CZS6
37	8	0.6	154	11	Q8R331
38	8	0.6	155	4	Q96E55
39	8	0.6	164	10	O82714
40	8	0.6	164	10	O43489
41	8	0.6	172	6	O95JR8
42	8	0.6	178	4	O9H3V1
43	8	0.6	186	10	O9FTI8
44	8	0.6	187	16	O8XW64
45	8	0.6	192	16	O8XGV9
46	8	0.6	192	16	O8XD01
47	8	0.6	198	16	O8Y2A58
48	8	0.6	198	16	O8Y5V5
49	8	0.6	199	16	O8XE89
50	8	0.6	204	10	Q38895
51	8	0.6	204	16	O9AB57
52	8	0.6	224	10	O93244
53	8	0.6	224	11	Q9JHE6
54	8	0.6	225	4	O9HC76
55	8	0.6	225	4	Q96FC6
56	8	0.6	225	4	O8TE24
57	8	0.6	225	6	O9BFM9
58	8	0.6	228	10	Q98S46
59	8	0.6	239	16	Q9KVL3
60	8	0.6	245	10	O8RYE9
61	8	0.6	248	4	O9H2W1
62	8	0.6	248	4	O8TE25
63	8	0.6	248	4	O8TBV7
64	8	0.6	249	4	O9H2N3
65	8	0.6	253	10	Q9M5A8
66	8	0.6	255	5	Q8T857
67	8	0.6	259	10	Q94BG5
68	8	0.6	272	2	Q02304
69	8	0.6	273	16	Q8YWX4
70	8	0.6	274	16	P75019
71	8	0.6	295	16	O8Y6Y4
72	8	0.6	304	2	Q933C0
73	8	0.6	309	12	O8U2B4
74	8	0.6	310	2	O85727

75	8	0.6	313	10	Q9C6H3	Q9c6h3 arabidopsis	c 148	8	0.6	971	2	Q9XC25	Q9xc25 bacillus an
c 76	8	0.6	322	16	Q9WXT4	Q9wxt4 thermotoga	149	8	0.6	1066	5	Q8T3T0	Q8t3t0 dictyosteli
c 77	8	0.6	346	12	Q91FG5	Q91fg5 chilo iride	c 150	8	0.6	1070	4	Q8WX12	Q8wx12 homo sapien
c 78	8	0.6	352	10	Q04792	Q04792 garcinia ma	151	8	0.6	1095	5	Q9U4G4	Q9u4g4 drosophilla
c 79	8	0.6	356	17	Q9HS47	Q9hs47 halobacteri	c 152	8	0.6	1118	10	Q93VL6	Q93vl6 phaseolus v
c 80	8	0.6	360	2	Q48888	Q48888 mycobacteri	c 153	8	0.6	1126	5	Q8T278	Q8t278 dictyosteli
c 81	8	0.6	360	16	Q97PV7	Q97pv7 streptococ	c 154	8	0.6	1133	10	Q93VS9	Q93vs9 phaseolus v
c 82	8	0.6	362	10	Q39402	Q39402 brassica ca	155	8	0.6	1152	10	Q9C8F4	Q9c8f4 arabidopsis
c 83	8	0.6	362	10	Q42561	Q42561 arabidopsis	c 156	8	0.6	1199	12	Q83859	Q83859 nilaparvata
c 84	8	0.6	369	16	Q9PM30	Q9pm30 campylobact	157	8	0.6	1322	4	Q8TB50	Q8te50 homo sapien
c 85	8	0.6	374	2	Q47467	Q47467 erwinia car	158	8	0.6	1510	2	Q92465	Q92465 corynebacte
c 86	8	0.6	381	16	Q98QV9	Q98qv9 mycoplasma	159	8	0.6	1514	4	Q96Q9A	Q96q9a homo sapien
c 87	8	0.6	382	16	Q98CP3	Q98cp3 rhizobium l	160	8	0.6	1561	11	Q924D2	Q924d2 mus musculu
c 88	8	0.6	383	16	Q8YE66	Q8ye66 bruceella me	c 161	8	0.6	1581	16	Q92C12	Q92ci2 rickettsia
c 89	8	0.6	389	11	Q8R3I9	Q8r3i9 mus musculu	c 162	8	0.6	1583	16	Q92GE9	Q92ge9 rickettsia
c 90	8	0.6	392	13	Q98UJ7	Q98uj7 gallus gall	c 163	8	0.6	1618	11	Q9QX19	Q9qx19 rattus norv
c 91	8	0.6	415	17	Q9HMQ7	Q9hmq7 halobacteri	c 164	8	0.6	1641	11	Q88528	Q88528 mus musculu
c 92	8	0.6	416	8	Q78234	Q78234 chrysosplen	c 165	8	0.6	1754	4	Q9U1W2	Q9u1w2 homo sapien
c 93	8	0.6	424	16	Q9K044	Q9k044 neisseria m	c 166	8	0.6	1894	11	P70206	P70206 mus musculu
c 94	8	0.6	424	16	Q9JV54	Q9jv54 neisseria m	c 167	8	0.6	1905	13	Q91823	Q91823 xenopus lae
c 95	8	0.6	427	16	Q8ZFN2	Q8zfn2 versinia pe	c 168	8	0.6	2023	5	Q96542	Q96542 drosophilla
c 96	8	0.6	429	12	Q9DWH7	Q9dwh7 rat cytomeg	c 169	8	0.6	2023	5	Q9V529	Q9v529 drosophilla
c 97	8	0.6	446	13	Q13006	Q13006 ictalurus p	c 170	8	0.6	2055	11	Q88938	Q88938 mus musculu
c 98	8	0.6	456	16	Q8UDF7	Q8udf7 agrobacteri	c 171	8	0.6	2102	5	Q18183	Q18183 caenorhabdi
c 99	8	0.6	460	16	Q8UA82	Q8ua82 agrobacteri	c 172	8	0.6	2778	5	Q9V9T6	Q9v9t6 drosophilla
c 100	8	0.6	464	5	Q95SH5	Q95sh5 drosophilla	173	8	0.6	3567	11	Q9ES77	Q9es77 mus musculu
101	8	0.6	464	5	Q9W2A0	Q9w2a0 drosophilla	c 174	8	0.6	3571	10	Q9SL27	Q9sl27 arabidopsis
102	8	0.6	478	16	Q9KV20	Q9kv20 vibrio chol	c 175	8	0.6	3574	10	Q9AUB4	Q9aub4 arabidopsis
c 103	8	0.6	480	13	Q13007	Q13007 ictalurus p	c 176	8	0.6	4060	12	Q91H28	Q91hz8 gill-associ
c 104	8	0.6	501	11	Q8VCW5	Q8vcw5 mus musculu	c 177	8	0.6	4924	3	Q9P7T1	Q9p7t1 schizosacch
c 105	8	0.6	508	8	Q78332	Q78332 perglularia	178	8	0.5	29	6	Q8WP11	Q8wp11 ateles belz
c 106	8	0.6	511	10	Q9XIH7	Q9xih7 arabidopsis	c 179	7	0.5	33	12	Q918E8	Q918e8 human cytom
c 107	8	0.6	511	10	Q94IH6	Q94ih6 arabidopsis	c 180	7	0.5	33	12	Q918E5	Q918e5 human cytom
c 108	8	0.6	518	10	Q94HA3	Q94ha3 oryza sativ	c 181	7	0.5	33	12	Q910Q5	Q910q5 human cytom
c 109	8	0.6	519	10	Q9STG8	Q9stg8 arabidopsis	182	7	0.5	43	2	Q46785	Q46785 escherichia
c 110	8	0.6	527	12	Q9GQ33	Q9gq33 grapevine l	183	7	0.5	43	2	Q47127	Q47127 escherichia
c 111	8	0.6	531	10	Q80735	Q80735 arabidopsis	184	7	0.5	43	2	Q57131	Q57131 escherichia
c 112	8	0.6	539	5	Q8SMW3	Q8smw3 drosophilla	185	7	0.5	45	2	Q9GF82	Q9gff8 gnetum gnem
c 113	8	0.6	559	16	Q8VRR7	Q8vrr7 anabaena sp	186	7	0.5	45	2	Q24683	Q24683 shigella bo
c 114	8	0.6	574	4	Q9H1G5	Q9h1g5 homo sapien	187	7	0.5	49	2	Q9MR6	Q9rm6 klebsiella
c 115	8	0.6	577	11	Q9D221	Q9d221 mus musculu	188	7	0.5	49	16	Q8U5B4	Q8u5b4 agrobacteri
c 116	8	0.6	595	2	Q34206	Q34206 pseudomonas	c 189	7	0.5	56	7	Q9MX88	Q9mx88 oreochromis
c 117	8	0.6	595	13	Q9DBE1	Q9dbec1 brachydanio	190	7	0.5	56	8	Q952Q2	Q952q2 diadema ant
c 118	8	0.6	595	16	Q9HT87	Q9ht87 pseudomonas	191	7	0.5	56	8	Q952Q0	Q952q0 diadema ant
c 119	8	0.6	598	10	Q9SSM0	Q9ssm0 arabidopsis	192	7	0.5	56	8	Q952P8	Q952p8 diadema ant
c 120	8	0.6	601	11	Q9CUT3	Q9cut3 mus musculu	193	7	0.5	56	8	Q952P7	Q952p7 diadema ant
c 121	8	0.6	628	16	Q8VHF6	Q8vhf6 bruceella me	194	7	0.5	56	8	Q952P5	Q952p5 diadema ant
c 122	8	0.6	630	3	Q13870	Q13870 schizosacch	195	7	0.5	56	8	Q952P3	Q952p3 diadema ant
c 123	8	0.6	653	15	Q87039	Q87039 simian foam	196	7	0.5	56	8	Q952P1	Q952p1 diadema ant
c 124	8	0.6	669	11	Q922H0	Q922h0 mus musculu	197	7	0.5	56	8	Q952P0	Q952p0 diadema ant
c 125	8	0.6	670	2	Q9RHQ0	Q9rhq0 pseudomonas	198	7	0.5	56	8	Q952N9	Q952n9 diadema ant
c 126	8	0.6	670	16	Q91596	Q91596 pseudomonas	199	7	0.5	56	8	Q952N7	Q952n7 diadema ant
c 127	8	0.6	674	11	Q60979	Q60979 mus musculu	200	7	0.5	56	8	Q952N5	Q952n5 diadema ant
c 128	8	0.6	680	2	Q9KH14	Q9kh14 mycoplasma	201	7	0.5	56	8	Q952M6	Q952m6 diadema ant
c 129	8	0.6	684	16	Q917B8	Q917b8 pseudomonas	202	7	0.5	56	8	Q952M3	Q952m3 diadema ant
c 130	8	0.6	694	10	Q9LE58	Q9le58 arabidopsis	203	7	0.5	56	8	Q952J4	Q952j4 diadema ant
c 131	8	0.6	717	10	Q9LXL0	Q9lxl0 arabidopsis	204	7	0.5	56	8	Q952I5	Q952i5 diadema ant
c 132	8	0.6	717	16	Q9PBZ8	Q9pbz8 xylella fas	205	7	0.5	56	8	Q952I2	Q952i2 diadema ant
c 133	8	0.6	724	3	Q8X0Y2	Q8x0y2 neurospora	206	7	0.5	56	8	Q952H6	Q952h6 diadema ant
c 134	8	0.6	748	10	Q9T0G5	Q9t0g5 arabidopsis	207	7	0.5	56	8	Q952H3	Q952h3 diadema ant
c 135	8	0.6	752	10	Q9FKZ9	Q9fkz9 arabidopsis	208	7	0.5	56	8	Q94FJ3	Q94fp3 diadema ant
c 136	8	0.6	754	5	Q8T769	Q8t769 brachiosteo	209	7	0.5	56	8	Q94N68	Q94n68 diadema ant
c 137	8	0.6	761	5	Q9VVL5	Q9vvl5 drosophilla	210	7	0.5	56	8	Q94N66	Q94n66 diadema ant
c 138	8	0.6	772	4	Q9Y468	Q9y468 homo sapien	211	7	0.5	56	8	Q94N51	Q94n51 diadema ant
c 139	8	0.6	792	5	Q9W2R3	Q9w2r3 drosophilla	212	7	0.5	56	8	Q94N50	Q94n50 diadema ant
c 140	8	0.6	793	3	Q9HFW2	Q9hfw2 ashbya goss	213	7	0.5	56	8	Q94P45	Q94p45 diadema ant
c 141	8	0.6	793	5	Q93325	Q93325 caenorhabdi	214	7	0.5	56	8	Q952N2	Q952n2 diadema ant
c 142	8	0.6	834	16	Q9PD71	Q9pd71 xylella fas	215	7	0.5	56	8	Q94N67	Q94n67 diadema ant
c 143	8	0.6	881	10	Q9SMX9	Q9smx9 arabidopsis	216	7	0.5	56	8	Q952P6	Q952p6 diadema ant
c 144	8	0.6	881	10	Q82651	Q82651 arabidopsis	217	7	0.5	56	8	Q94N49	Q94n49 diadema ant
c 145	8	0.6	887	5	Q8STF6	Q8stf6 caenorhabdi	218	7	0.5	56	8	Q8WFV4	Q8wfv4 diadema ant
c 146	8	0.6	907	16	Q9L248	Q9l248 streptomyce	219	7	0.5	56	8	Q8WF55	Q8wff5 diadema mex
c 147	8	0.6	925	2	Q9RQ00	Q9rq00 bacillus an	220	7	0.5	56	8	Q8WFO9	Q8wff9 diadema pau

221	7	0.5	56	8	Q8WFO5	Q8wfg5 diadema pau	c 294	7	0.5	114	5	Q95Q94	Q95q94 caenorhabdi
222	7	0.5	56	8	Q8WFO4	Q8wfg4 diadema pau	c 295	7	0.5	114	16	Q99TS7	Q99ts7 staphylococ
223	7	0.5	56	8	Q8WFO2	Q8wfg2 diadema pau	c 296	7	0.5	115	12	Q9Q2N1	Q9q2n1 human eryth
224	7	0.5	56	8	Q8WFO1	Q8wfg1 diadema pau	c 297	7	0.5	116	17	Q97OT8	Q97ot8 sulfolobus
225	7	0.5	56	8	Q8WFP1	Q8wfp1 diadema pau	c 298	7	0.5	117	16	Q9UXP8	Q9xup8 neisseria m
226	7	0.5	56	8	Q8WFP0	Q8wfp0 diadema pau	c 299	7	0.5	117	16	Q9JW74	Q9jw74 neisseria m
227	7	0.5	56	8	Q8WFN8	Q8wfn8 diadema pau	c 300	7	0.5	117	16	Q97QR9	Q97qr9 streptococ
228	7	0.5	56	8	Q8WFN7	Q8wfn7 diadema pau	c 301	7	0.5	118	2	Q9AQG0	Q9aqg0 acidovorax
229	7	0.5	56	8	Q8WFM2	Q8wfm2 diadema sav	c 302	7	0.5	118	4	Q9NQF6	Q9nqf6 homo sapien
230	7	0.5	56	8	Q8WFM4	Q8wfm4 diadema ant	c 303	7	0.5	118	5	Q95X43	Q95x43 caenorhabdi
231	7	0.5	56	8	Q8W7R7	Q8w7r7 diadema sav	c 304	7	0.5	118	5	Q9UIK5	Q9uik5 drosophila
232	7	0.5	56	8	Q8W7L8	Q8w7l8 diadema mex	c 305	7	0.5	118	8	Q9ME35	Q9mep35 oxytoma hay
233	7	0.5	56	8	Q8W7L6	Q8w7l6 diadema sav	c 306	7	0.5	118	8	Q9MEP0	Q9mep0 oxytoma hay
234	7	0.5	56	8	Q8W7C9	Q8w7c9 diadema pau	c 307	7	0.5	118	10	Q49947	Q49947 solanum tub
235	7	0.5	56	8	Q8W7C8	Q8w7c8 diadema sav	c 308	7	0.5	118	16	Q8R8Y7	Q8r8y7 thermoanaer
236	7	0.5	65	6	Q97832	Q97832 canis famil	c 309	7	0.5	119	5	Q9V401	Q9v401 drosophila
237	7	0.5	65	7	Q77940	Q77940 sus scrofa	c 310	7	0.5	119	5	Q9V8F3	Q9v8f3 drosophila
238	7	0.5	68	10	Q9S967	Q9s967 medicago sa	c 311	7	0.5	119	10	Q9ZSF5	Q9zsf5 lycium ande
239	7	0.5	69	2	Q57052	Q57052 staphylococ	c 312	7	0.5	120	4	Q75591	Q75591 homo sapien
240	7	0.5	70	2	Q9EXE4	Q9exe4 lactobacilli	c 313	7	0.5	120	4	Q13542	Q13542 homo sapien
241	7	0.5	70	10	Q8S863	Q8s863 oryza sativ	c 314	7	0.5	120	16	Q8Y379	Q8y379 ralstonia s
242	7	0.5	71	10	Q8S8P7	Q8s8p7 arabidopsis	c 315	7	0.5	121	12	Q9YRJ5	Q9yrl5 human cytom
243	7	0.5	71	16	Q9JQW9	Q9jqw9 neisseria m	c 316	7	0.5	121	17	Q97VK2	Q97vk2 sulfolobus
244	7	0.5	73	8	Q20453	Q20453 fugu rubrip	c 317	7	0.5	122	12	Q9YRJ1	Q9yrl1 human cytom
245	7	0.5	73	8	Q9XQR3	Q9xqr3 pisum sativ	c 318	7	0.5	122	16	Q8YAW3	Q8yaw3 bruceella me
246	7	0.5	75	16	Q92NY7	Q92ny7 rhizobium m	c 319	7	0.5	123	2	Q32992	Q32992 mycobacteri
247	7	0.5	79	12	Q68100	Q68100 human cytom	c 320	7	0.5	123	8	Q9MJM2	Q9mjm2 limnodynast
248	7	0.5	79	16	Q8XXZ3	Q8xxz3 ralstonia s	c 321	7	0.5	123	11	Q9CZ38	Q9cz38 mus musculu
249	7	0.5	80	16	Q9CA50	Q9ca50 arabidopsis	c 322	7	0.5	123	12	Q9BDJ9	Q9bdj9 cherry gree
250	7	0.5	80	16	Q9PP50	Q9pp50 campylobact	c 323	7	0.5	124	5	Q95RN4	Q95rn4 drosophila
251	7	0.5	82	15	Q9YKY7	Q9yky7 human immun	c 324	7	0.5	124	10	Q92PV4	Q92pv4 arabidopsis
252	7	0.5	83	2	Q9RDW4	Q9rdw4 lactobacilli	c 325	7	0.5	124	17	Q976C8	Q976c8 sulfolobus
253	7	0.5	83	2	Q8VRG2	Q8vrg2 lactobacilli	c 326	7	0.5	124	17	Q26783	Q26783 methanobact
254	7	0.5	85	11	Q8S095	Q8s095 mus musculu	c 327	7	0.5	125	6	Q95K14	Q95k14 macaca fasc
255	7	0.5	87	5	Q94787	Q94787 trypanosoma	c 328	7	0.5	125	8	Q9XN44	Q9xn44 tupinambis
256	7	0.5	88	17	Q970B1	Q970b1 sulfolobus	c 329	7	0.5	125	8	Q9TFC1	Q9tfcl tupinambis
257	7	0.5	89	5	Q9V542	Q9v542 drosophila	c 330	7	0.5	125	8	Q9TFC0	Q9tfco tupinambis
258	7	0.5	90	8	Q953X0	Q953x0 terebratali	c 331	7	0.5	125	8	Q9XN43	Q9xn43 cnemidophor
259	7	0.5	92	16	Q9JXN2	Q9jxn2 neisseria m	c 332	7	0.5	125	8	Q9XN42	Q9xn42 ameiva amei
260	7	0.5	95	10	Q8S2N1	Q8s2n1 oryza sativ	c 333	7	0.5	125	8	Q9XN41	Q9xn41 kentropox v
261	7	0.5	95	16	Q9K102	Q9k102 neisseria m	c 334	7	0.5	125	8	Q9XN40	Q9xn40 tupinambis
262	7	0.5	96	16	Q8RHS9	Q8rhs9 fusobacteri	c 335	7	0.5	125	8	Q9TFB9	Q9tfb9 tupinambis
263	7	0.5	98	12	Q9YR19	Q9yri9 human cytom	c 336	7	0.5	125	8	Q9XK25	Q9xk25 tupinambis
264	7	0.5	100	3	Q9B681	Q9b681 saccharomyc	c 337	7	0.5	125	8	Q9XJZ9	Q9xjz9 tupinambis
265	7	0.5	100	12	Q9YRJ4	Q9yrl4 human cytom	c 338	7	0.5	125	8	Q9XJZ8	Q9xjz8 tupinambis
266	7	0.5	100	16	Q9JW91	Q9jw91 neisseria m	c 339	7	0.5	125	8	Q9T3U5	Q9t3u5 tupinambis
267	7	0.5	100	17	Q9TDP9	Q9tdp9 aeropyrum p	c 340	7	0.5	125	8	Q9T3M0	Q9t3m0 tupinambis
268	7	0.5	101	12	Q9YR17	Q9yrl7 human cytom	c 341	7	0.5	125	17	Q9YCW4	Q9ycw4 aeropyrum p
269	7	0.5	101	15	Q9YKR3	Q9ykr3 human immun	c 342	7	0.5	126	16	Q8YKM8	Q8ykm8 anabaena sp
270	7	0.5	102	2	Q50225	Q50225 thiobacillu	c 343	7	0.5	126	16	Q8YKD6	Q8ykd6 anabaena sp
271	7	0.5	102	16	Q9RTH8	Q9rth8 deinococcus	c 344	7	0.5	128	2	Q8RLQ4	Q8rlq4 providencia
272	7	0.5	103	15	Q8UT31	Q8ut31 human immun	c 345	7	0.5	128	3	Q9HGT1	Q9hgt1 coccidioidi
273	7	0.5	104	2	Q45350	Q45350 beta proteo	c 346	7	0.5	128	5	Q9VME4	Q9vme4 drosophila
274	7	0.5	104	5	Q8SVF6	Q8svf6 encephalito	c 347	7	0.5	128	8	Q9MJN8	Q9mjn8 limnodynast
275	7	0.5	105	5	Q9VPE5	Q9vpe5 drosophila	c 348	7	0.5	128	12	Q68084	Q68084 human cytom
276	7	0.5	105	5	Q8T3M1	Q8t3m1 drosophila	c 349	7	0.5	129	2	Q52356	Q52356 mycoplasma
277	7	0.5	105	11	Q9D4U9	Q9d4u9 mus musculu	c 350	7	0.5	129	4	Q9BXM8	Q9bxm8 homo sapien
278	7	0.5	105	11	Q9S774	Q9s774 mus musculu	c 351	7	0.5	130	4	Q9BSB0	Q9bsb0 homo sapien
279	7	0.5	105	12	Q9YRJ6	Q9yrl6 human cytom	c 352	7	0.5	130	15	Q76870	Q76870 human immun
280	7	0.5	106	12	Q9WKR3	Q9wkr3 human cytom	c 353	7	0.5	130	15	Q76871	Q76871 human immun
281	7	0.5	106	12	Q88468	Q88468 strawberry	c 354	7	0.5	130	15	Q76872	Q76872 human immun
282	7	0.5	107	16	Q9JXU1	Q9jxu1 neisseria m	c 355	7	0.5	130	15	Q76873	Q76873 human immun
283	7	0.5	109	6	Q28118	Q28118 bos taurus	c 356	7	0.5	130	15	Q76874	Q76874 human immun
284	7	0.5	110	10	Q94HZ7	Q94hz7 oryza sativ	c 357	7	0.5	130	15	Q76877	Q76877 human immun
285	7	0.5	110	17	Q58195	Q58195 pyrococcus	c 358	7	0.5	130	15	Q76878	Q76878 human immun
286	7	0.5	111	5	Q96509	Q96509 caenorhabdi	c 359	7	0.5	130	15	Q76879	Q76879 human immun
287	7	0.5	111	11	Q88435	Q88435 rattus norv	c 360	7	0.5	130	15	Q76880	Q76880 human immun
288	7	0.5	111	12	Q9YRJ8	Q9yrl8 human cytom	c 361	7	0.5	130	15	Q76881	Q76881 human immun
289	7	0.5	111	16	Q9RTG6	Q9rtg6 deinococcus	c 362	7	0.5	130	15	Q76900	Q76900 human immun
290	7	0.5	112	11	Q8VDH3	Q8vdh3 mus musculu	c 363	7	0.5	130	15	Q76901	Q76901 human immun
291	7	0.5	112	12	Q9YRJ7	Q9yrl7 human cytom	c 364	7	0.5	130	15	Q76902	Q76902 human immun
292	7	0.5	113	16	Q98MW1	Q98mw1 rhizobium l	c 365	7	0.5	130	15	Q76903	Q76903 human immun
293	7	0.5	113	16	Q98K14	Q98k14 rhizobium l	c 366	7	0.5	130	15	Q76904	Q76904 human immun

c 367	7	0.5	130	15	Q76905	Q76905 human immun	c 440	7	0.5	145	13	Q9DG79	Q9dg79 ictalurus p
c 368	7	0.5	130	15	Q76907	Q76907 human immun	c 441	7	0.5	146	15	Q90BC4	Q90bc4 human immun
c 369	7	0.5	130	15	Q76908	Q76908 human immun	442	7	0.5	146	16	P94515	P94515 bacillus su
c 370	7	0.5	130	15	Q76909	Q76909 human immun	443	7	0.5	146	16	Q8UBR4	Q8ubr4 agrobacteri
c 371	7	0.5	130	15	Q76910	Q76910 human immun	444	7	0.5	146	16	Q8R708	Q8r708 thermoanaer
c 372	7	0.5	130	15	Q77049	Q77049 human immun	c 445	7	0.5	147	5	Q9GYM0	Q9gym0 caenorhabdi
c 373	7	0.5	130	15	Q77050	Q77050 human immun	c 446	7	0.5	148	4	Q9UWJ4	Q9umj4 homo sapien
c 374	7	0.5	130	15	Q77052	Q77052 human immun	c 447	7	0.5	148	10	Q49179	Q49179 oryza sativ
c 375	7	0.5	130	15	Q77055	Q77055 human immun	448	7	0.5	148	10	Q8VXY1	Q8vxy1 arabidopsis
c 376	7	0.5	130	15	Q77057	Q77057 human immun	449	7	0.5	148	10	Q8VXY1	Q8vxy1 arabidopsis
c 377	7	0.5	130	15	Q77058	Q77058 human immun	c 450	7	0.5	149	6	Q9TVB4	Q9tvb4 canis famil
c 378	7	0.5	130	15	Q77088	Q77088 human immun	451	7	0.5	150	8	Q9MLH3	Q9mlh3 paragonimus
c 379	7	0.5	130	15	Q77089	Q77089 human immun	c 452	7	0.5	150	8	Q9MLH3	Q9mlh3 paragonimus
380	7	0.5	131	8	Q9MJM4	Q9mj4 limnodynast	c 453	7	0.5	150	16	Q55491	Q55491 synechocyst
381	7	0.5	131	10	Q9XES9	Q9xes9 glycine max	454	7	0.5	150	16	Q9G011	Q9g011 rhizobium l
382	7	0.5	131	10	Q9LKU5	Q9lku5 arabidopsis	c 455	7	0.5	150	16	Q92IJ6	Q92ij6 rickettsia
383	7	0.5	132	8	Q9MJN3	Q9mj3 limnodynast	456	7	0.5	151	2	Q68739	Q68739 yersinia pe
384	7	0.5	134	9	Q34035	Q34035 streptococc	457	7	0.5	151	10	Q9FWE5	Q9fwe5 oryza sativ
c 385	7	0.5	134	10	Q94GZ2	Q94gz2 oryza sativ	c 458	7	0.5	151	16	Q84177	Q84177 chlamydia t
c 386	7	0.5	134	16	Q9RSP5	Q9rsp5 deinococcus	c 459	7	0.5	151	16	Q9CGK9	Q9cgk9 lactococcus
c 387	7	0.5	135	5	Q17248	Q17248 caenorhabdi	460	7	0.5	152	5	Q95ZL9	Q95zl9 caenorhabdi
388	7	0.5	135	8	Q9MJP7	Q9mjp7 crinia dese	c 461	7	0.5	152	16	Q9ABX2	Q9abx2 caulobacter
389	7	0.5	135	8	Q9MJP6	Q9mjp6 crinia pari	462	7	0.5	153	5	Q9UIK3	Q9ulk3 drosophila
c 390	7	0.5	135	10	Q9FXQ7	Q9fxq7 zea mays (m	463	7	0.5	153	5	Q9VFI5	Q9vfi5 drosophila
c 391	7	0.5	135	16	Q9RJZ5	Q9rjz5 streptomyce	c 464	7	0.5	154	5	Q8T972	Q8t972 drosophila
392	7	0.5	136	8	Q9MEI9	Q9mei9 limnodynast	465	7	0.5	154	16	Q8UAI4	Q8uai4 agrobacteri
393	7	0.5	136	8	Q9MDW5	Q9mdw5 limnodynast	466	7	0.5	154	16	Q9X8V9	Q9x8v9 aeropyrum p
394	7	0.5	136	8	Q9MDU2	Q9mdu2 limnodynast	c 467	7	0.5	155	5	Q9TXS7	Q9txs7 caenorhabdi
395	7	0.5	136	8	Q9MDS3	Q9mds3 limnodynast	c 468	7	0.5	156	10	Q41718	Q41718 zea diplope
396	7	0.5	136	8	Q9MDS2	Q9mds2 limnodynast	c 469	7	0.5	156	10	Q93YF6	Q93yf6 nicotiana s
397	7	0.5	136	8	Q9MDM3	Q9mdm3 limnodynast	470	7	0.5	156	15	Q92813	Q92813 bovine leuk
398	7	0.5	136	8	Q9MJN9	Q9mj9 limnodynast	c 471	7	0.5	156	16	Q9A9N5	Q9a9n5 caulobacter
399	7	0.5	136	8	Q9MJP5	Q9mjp5 adelotus br	472	7	0.5	157	5	Q9WLK2	Q9wlk2 drosophila
400	7	0.5	136	8	Q9MJP4	Q9mjp4 megistoloti	c 473	7	0.5	157	16	Q9P1P3	Q9p1p3 campylobact
401	7	0.5	136	8	Q9MJP1	Q9mjp1 limnodynast	474	7	0.5	157	16	Q9P9P4	Q9p9p4 xylella fas
402	7	0.5	136	8	Q9MJN7	Q9mj7 limnodynast	475	7	0.5	158	10	Q941E5	Q941e5 arabidopsis
403	7	0.5	136	8	Q9MJN6	Q9mj6 limnodynast	476	7	0.5	158	12	Q56249	Q56249 tobacco rat
404	7	0.5	136	8	Q9MJN5	Q9mj5 limnodynast	c 477	7	0.5	159	10	Q22690	Q22690 arabidopsis
405	7	0.5	136	8	Q9MJN4	Q9mj4 limnodynast	c 478	7	0.5	159	17	Q97YG6	Q97yg6 sulfolobus
406	7	0.5	136	8	Q9MJN2	Q9mj2 limnodynast	c 479	7	0.5	160	8	Q8WAG1	Q8wag1 sphaerospir
407	7	0.5	136	8	Q9MJN1	Q9mj1 limnodynast	c 480	7	0.5	160	8	Q8WAG0	Q8wag0 sphaerospir
408	7	0.5	136	8	Q9MJN0	Q9mj0 limnodynast	c 481	7	0.5	160	8	Q8WAF9	Q8waf9 sphaerospir
409	7	0.5	136	8	Q9MJM9	Q9mj9 limnodynast	c 482	7	0.5	160	16	Q8WAF8	Q8waf8 sphaerospir
410	7	0.5	136	8	Q9MJM8	Q9mj8 limnodynast	c 483	7	0.5	160	16	Q9KM07	Q9km07 vibrio chol
411	7	0.5	136	8	Q9MJM7	Q9mj7 limnodynast	c 484	7	0.5	160	17	Q58511	Q58511 pyrococcus
412	7	0.5	136	8	Q9MJM6	Q9mj6 limnodynast	485	7	0.5	160	17	Q97WV2	Q97wv2 sulfolobus
413	7	0.5	136	8	Q9MJM5	Q9mj5 limnodynast	c 486	7	0.5	161	4	Q9NWZ0	Q9nwz0 homo sapien
414	7	0.5	136	8	Q9MJM3	Q9mj3 limnodynast	c 487	7	0.5	161	4	Q9NT07	Q9nty07 homo sapien
415	7	0.5	136	8	Q9MJM1	Q9mj1 limnodynast	c 488	7	0.5	161	4	Q9BVL1	Q9bvl1 homo sapien
c 416	7	0.5	136	16	Q9JDM3	Q9jum3 neisseria m	489	7	0.5	161	16	Q9PCH9	Q9pch9 xylella fas
c 417	7	0.5	137	12	Q9E2L6	Q9e2l6 hepatitis c	490	7	0.5	161	17	Q26735	Q26735 methanobact
418	7	0.5	137	12	Q919L8	Q919l8 culex nigri	c 491	7	0.5	162	5	Q23137	Q23137 caenorhabdi
419	7	0.5	137	16	Q8ZFB8	Q8zfb8 yersinia pe	c 492	7	0.5	162	8	Q95AX6	Q95ax6 cyclanthus
c 420	7	0.5	138	11	Q9CUP0	Q9cup0 mus musculus	c 493	7	0.5	162	8	Q95AX0	Q95ax0 pandanus ut
421	7	0.5	138	12	Q9YR15	Q9yr15 human cytom	494	7	0.5	162	16	Q8XQU9	Q8xqu9 raistonia s
c 422	7	0.5	138	15	Q41391	Q41391 human immun	c 495	7	0.5	163	17	Q9YGL1	Q9ygl1 aeropyrum p
423	7	0.5	138	16	Q9RX00	Q9rx00 deinococcus	c 496	7	0.5	164	2	Q9EUJ9	Q9euj9 salmonella
c 424	7	0.5	138	16	Q9A9G5	Q9a9g5 caulobacter	c 497	7	0.5	164	2	Q70017	Q70017 shigella dy
425	7	0.5	138	16	Q8UGI0	Q8ug10 agrobacteri	c 498	7	0.5	164	4	Q96QK6	Q96qk6 homo sapien
426	7	0.5	139	10	Q9LNI6	Q9lni6 arabidopsis	c 499	7	0.5	164	10	Q8WZ24	Q8wz24 oryza sativ
427	7	0.5	139	12	Q9YRJ2	Q9yrj2 human cytom	c 500	7	0.5	164	16	Q8X5N5	Q8x5n5 escherichia
c 428	7	0.5	139	13	Q9DG76	Q9dg76 ictalurus p	c 501	7	0.5	164	16	Q8VKH4	Q8vkh4 mycobacteri
429	7	0.5	139	16	Q8Z5Q2	Q8z5q2 salmonella	502	7	0.5	165	15	Q91G02	Q91g02 bovine leuk
430	7	0.5	140	10	Q8S1H1	Q8s1h1 oryza sativ	c 503	7	0.5	165	16	Q9Z1H9	Q9z1h9 clostridium
431	7	0.5	140	12	Q9QUJ35	Q9quj35 ttv-like ml	c 504	7	0.5	165	16	Q9X9V2	Q9x9v2 streptomyce
c 432	7	0.5	142	6	Q9GKN3	Q9gkn3 bos taurus	505	7	0.5	166	16	Q9A3E6	Q9a3e6 caulobacter
433	7	0.5	143	5	Q9N7Y0	Q9n7y0 leishmania	506	7	0.5	166	16	Q9ADM6	Q9adm6 streptomyce
434	7	0.5	143	12	Q90189	Q90189 bombyx mori	c 507	7	0.5	168	2	Q93IK3	Q93ik3 vibrio sp.
c 435	7	0.5	143	16	Q8YHU9	Q8yhu9 brucella me	c 508	7	0.5	168	16	Q92SI5	Q92si5 rhizobium m
436	7	0.5	144	5	Q9GZ13	Q9gz13 caenorhabdi	c 509	7	0.5	169	8	Q94S76	Q94s76 zenopsis ne
c 437	7	0.5	145	2	Q9AP12	Q9ap12 legionella	c 510	7	0.5	169	12	Q912C6	Q912c6 human eryth
c 438	7	0.5	145	5	Q9VTF3	Q9vtf3 drosophila	c 511	7	0.5	170	2	Q9KKI9	Q9kki9 yersinia en
c 439	7	0.5	145	11	Q9D6U0	Q9d6u0 mus musculus	512	7	0.5	170	16	Q8R6X4	Q8r6x4 thermoanaer

513	7	0.5	171	4	O00486	O00486 homo sapien	586	7	0.5	185	8	Q94Q98
c 514	7	0.5	171	16	O8UIR4	O8uir4 agrobacteri	587	7	0.5	185	8	Q94Q98
c 515	7	0.5	172	5	O61948	O61948 caenorhabdi	588	7	0.5	185	8	Q94Q64
c 516	7	0.5	172	8	Q966Q2	Q966q2 diplophos t	589	7	0.5	185	8	Q94PI0
c 517	7	0.5	173	8	Q966N9	Q966n9 polymixia j	590	7	0.5	185	8	Q94NN7
c 518	7	0.5	173	8	Q96B9U4	Q96b9u4 autopus jep	591	7	0.5	185	8	Q94NM0
c 519	7	0.5	173	8	Q94YR3	Q94yr3 engraulis j	592	7	0.5	185	8	Q94NG7
c 520	7	0.5	173	8	Q94T69	Q94t69 lampris gut	593	7	0.5	185	8	Q94NG6
c 521	7	0.5	173	8	Q94T35	Q94t35 polymixia l	594	7	0.5	185	8	Q94N10
c 522	7	0.5	173	8	Q94SK9	Q94sk9 zeus faber	595	7	0.5	185	16	Q9CHS8
c 523	7	0.5	173	8	Q94SF7	Q94sf7 monopteris	c 596	7	0.5	186	4	Q96ST6
c 524	7	0.5	174	2	Q9EXY7	Q9eyx7 klebsiella	597	7	0.5	186	8	Q9MIX5
c 525	7	0.5	174	2	Q8RTW3	Q8rtw3 uncultured	598	7	0.5	186	8	Q956V1
c 526	7	0.5	174	4	Q9NSU8	Q9nsu8 homo sapien	599	7	0.5	186	10	Q94IM0
c 527	7	0.5	174	5	Q9NGY3	Q9ngy3 difofilaria	c 600	7	0.5	186	16	Q9PB72
c 528	7	0.5	174	5	O76968	O76968 podocoryne	601	7	0.5	186	16	Q8RHL3
c 529	7	0.5	174	11	Q9CTX3	Q9ctx3 mus musculus	c 602	7	0.5	186	17	Q27711
c 530	7	0.5	175	5	Q23646	Q23646 caenorhabdi	c 603	7	0.5	187	10	O24352
c 531	7	0.5	175	11	Q9D623	Q9d623 mus musculus	604	7	0.5	187	10	O24020
c 532	7	0.5	175	11	Q9D6W5	Q9d6w5 mus musculus	605	7	0.5	188	5	Q9W012
c 533	7	0.5	175	17	O8TP57	O8tp57 methanosarc	c 606	7	0.5	188	6	O8WMD0
c 534	7	0.5	176	6	P79329	P79329 callitrix	607	7	0.5	188	10	Q9MAV1
c 535	7	0.5	176	16	Q9ABZ9	Q9abz9 caulobacter	c 608	7	0.5	188	16	Q9A3H5
c 536	7	0.5	176	16	O8RI01	O8ri01 fusobacteri	c 609	7	0.5	189	2	Q54918
c 537	7	0.5	177	4	Q96IM9	Q96im9 homo sapien	c 610	7	0.5	189	4	Q96GS0
c 538	7	0.5	177	8	Q956T0	Q956t0 limnodynast	c 611	7	0.5	189	16	Q985I8
c 539	7	0.5	177	12	O8OLJ4	O8olj4 mamestra co	612	7	0.5	189	16	O8Z614
c 540	7	0.5	177	17	Q9YCX1	Q9ycx1 aeropyrum p	613	7	0.5	189	16	O31771
c 541	7	0.5	178	2	Q54933	Q54933 streptococc	614	7	0.5	190	4	Q8TBP5
c 542	7	0.5	178	8	Q956U9	Q956u9 limnodynast	c 615	7	0.5	190	5	Q9V5P2
c 543	7	0.5	178	10	O48681	O48681 arabidopsis	c 616	7	0.5	190	6	O8WNM1
c 544	7	0.5	178	10	O8RVQ9	O8rvq9 arabidopsis	c 617	7	0.5	190	6	O8WNM0
c 545	7	0.5	178	11	Q9DA50	Q9da50 mus musculus	c 618	7	0.5	190	8	Q9B3V2
c 546	7	0.5	178	16	Q9RT80	Q9rt80 deinococcus	619	7	0.5	190	8	Q9B3U4
c 547	7	0.5	178	16	Q9HT17	Q9ht17 pseudomonas	620	7	0.5	190	8	Q8WC80
c 548	7	0.5	179	10	Q9C571	Q9c571 arabidopsis	621	7	0.5	190	11	Q9D3L0
c 549	7	0.5	179	16	Q99SF2	Q99sf2 staphylococ	622	7	0.5	190	11	Q9D238
c 550	7	0.5	180	5	Q9Q968	Q9q968 caenorhabdi	c 623	7	0.5	190	11	Q9D052
c 551	7	0.5	180	5	O18997	O18997 caenorhabdi	624	7	0.5	190	11	O9CYF8
c 552	7	0.5	180	8	Q956U2	Q956u2 limnodynast	625	7	0.5	190	11	Q8VDR1
c 553	7	0.5	180	16	Q9KMQ7	Q9kmq7 vibrio chol	626	7	0.5	190	12	Q90149
c 554	7	0.5	180	16	Q9CFU5	Q9cfu5 lactococcus	627	7	0.5	191	2	O51793
c 555	7	0.5	180	16	O8G314	O8g314 mycobacteri	628	7	0.5	191	4	Q96MY2
c 556	7	0.5	180	16	O8UJ07	O8uj07 agrobacteri	629	7	0.5	191	4	Q9B3U8
c 557	7	0.5	180	16	O8UID9	O8uid9 agrobacteri	630	7	0.5	192	8	O8WC81
c 558	7	0.5	181	5	O9GZ14	Q9gz14 caenorhabdi	c 631	7	0.5	192	11	Q8R3W8
c 559	7	0.5	181	8	Q956T2	Q956t2 limnodynast	632	7	0.5	192	16	Q912L9
c 560	7	0.5	181	10	Q9FGF1	Q9fgf1 arabidopsis	c 633	7	0.5	193	16	Q9JWK5
c 561	7	0.5	181	10	Q9LYH4	Q9lyh4 arabidopsis	634	7	0.5	193	16	Q9V827
c 562	7	0.5	182	5	O20866	O20866 caenorhabdi	635	7	0.5	194	8	O8WC83
c 563	7	0.5	182	5	O9VCG7	O9vcg7 drosophila	636	7	0.5	194	8	O8WC82
c 564	7	0.5	182	8	Q956U7	Q956u7 limnodynast	c 637	7	0.5	194	10	O8VXC4
c 565	7	0.5	182	8	Q956T9	Q956t9 limnodynast	c 638	7	0.5	195	4	Q9H7R4
c 566	7	0.5	182	8	Q956T1	Q956t1 limnodynast	c 639	7	0.5	195	5	Q9V9Z1
c 567	7	0.5	182	10	Q93ZG8	Q93zg8 arabidopsis	c 640	7	0.5	195	16	Q9RWJ5
c 568	7	0.5	182	11	Q9ER90	Q9er90 mus musculus	c 641	7	0.5	196	11	O8R1H7
c 569	7	0.5	182	16	O9I539	Q9i539 pseudomonas	642	7	0.5	197	8	O8WC84
c 570	7	0.5	183	8	Q956U6	Q956u6 limnodynast	643	7	0.5	197	10	Q9S9P9
c 571	7	0.5	184	8	Q956T7	Q956t7 limnodynast	c 644	7	0.5	197	16	O06983
c 572	7	0.5	185	8	Q956V0	Q956v0 limnodynast	645	7	0.5	198	8	O8WCC5
c 573	7	0.5	185	8	Q956G08	Q956g8 limnodynast	c 651	7	0.5	198	10	Q9M844
c 574	7	0.5	185	8	Q956U8	Q956u8 limnodynast	646	7	0.5	197	16	Q9JZK3
c 575	7	0.5	185	8	Q956U3	Q956u3 limnodynast	647	7	0.5	198	2	Q9F419
c 576	7	0.5	185	8	Q956U3	Q956u3 limnodynast	648	7	0.5	198	4	Q96GY7
c 577	7	0.5	185	8	Q956U1	Q956u1 limnodynast	c 649	7	0.5	198	5	O45186
c 578	7	0.5	185	8	Q956G0	Q956g0 limnodynast	650	7	0.5	198	8	O8WCC5
c 579	7	0.5	185	8	Q956T8	Q956t8 limnodynast	c 651	7	0.5	198	10	Q9M844
c 580	7	0.5	185	8	Q956T6	Q956t6 limnodynast	652	7	0.5	199	2	Q9RWF1
c 581	7	0.5	185	8	Q956T5	Q956t5 limnodynast	653	7	0.5	200	4	O00485
c 582	7	0.5	185	8	Q956T4	Q956t4 limnodynast	654	7	0.5	200	8	Q9B3U3
c 583	7	0.5	185	8	Q956T3	Q956t3 limnodynast	655	7	0.5	200	8	Q9B3U1
c 584	7	0.5	185	8	Q956S9	Q956s9 limnodynast	c 656	7	0.5	200	8	Q9B3U1
c 585	7	0.5	185	8	Q956S8	Q956s8 limnodynast	c 657	7	0.5	200	8	Q959Q7
c 586	7	0.5	185	8	Q94Q99	Q94q99 limnodynast	c 658	7	0.5	200	8	Q959Q3
c 587	7	0.5	185	8	Q94Q99	Q94q99 limnodynast	c 659	7	0.5	200	16	Q99TE4

659	7	0.5	200	16	Q8YBB1	Q8YBB1 brucella me	c 732	7	0.5	211	16	Q98410	Q98410 rhizobium l
c 660	7	0.5	201	10	P93539	P93539 silene prat	733	7	0.5	211	17	Q9YCC2	Q9YCC2 aeropyrum p
c 661	7	0.5	201	16	Q9AD42	Q9AD42 streptomyce	c 734	7	0.5	211	17	Q8TLT7	Q8TLT7 methanosarc
662	7	0.5	202	8	Q8WC85	Q8WC85 neusticurus	735	7	0.5	212	8	Q9B0K3	Q9B0K3 limnodynast
c 663	7	0.5	202	11	Q9DAP7	Q9DAP7 mus musculus	c 736	7	0.5	212	12	Q41980	Q41980 murid herpe
c 664	7	0.5	202	16	Q9AB41	Q9AB41 caulobacter	737	7	0.5	214	5	Q8WSH4	Q8WSH4 ancylostoma
665	7	0.5	203	8	Q9B3V3	Q9B3V3 plethodon f	738	7	0.5	214	8	Q9GCB6	Q9GCB6 litoria cae
666	7	0.5	203	8	Q8WC88	Q8WC88 anotosaura	739	7	0.5	214	8	Q9GCB5	Q9GCB5 litoria cit
667	7	0.5	203	8	Q8WC87	Q8WC87 calyptommat	740	7	0.5	214	8	Q9GCB4	Q9GCB4 litoria cit
668	7	0.5	203	8	Q8WC86	Q8WC86 micrablepha	741	7	0.5	214	8	Q9GCB3	Q9GCB3 litoria sub
c 669	7	0.5	203	11	Q9CV13	Q9CV13 mus musculus	742	7	0.5	214	8	Q9GCB2	Q9GCB2 litoria sub
670	7	0.5	204	2	Q93IH4	Q93IH4 wolinnella s	743	7	0.5	214	8	Q9GCB1	Q9GCB1 litoria sp.
c 671	7	0.5	204	10	Q23660	Q23660 arabidopsis	744	7	0.5	214	8	Q9GCB0	Q9GCB0 litoria spe
c 672	7	0.5	204	11	Q9QX18	Q9QX18 rattus norv	745	7	0.5	214	8	Q9GCA9	Q9GCA9 litoria nud
c 673	7	0.5	204	12	Q66087	Q66087 canine herp	746	7	0.5	214	8	Q9GCA8	Q9GCA8 litoria phy
c 674	7	0.5	204	16	Q8YGU3	Q8YGU3 brucella me	747	7	0.5	214	8	Q9GCA7	Q9GCA7 litoria pea
c 675	7	0.5	205	2	Q51712	Q51712 paracoccus	748	7	0.5	214	8	Q9GCA6	Q9GCA6 litoria pea
676	7	0.5	205	8	Q9B3V0	Q9B3V0 plethodon j	749	7	0.5	214	8	Q9B0J7	Q9B0J7 limnodynast
677	7	0.5	205	8	Q8WC89	Q8WC89 eclepous g	750	7	0.5	214	8	Q952E2	Q952E2 lineatriton
678	7	0.5	206	5	Q9NM09	Q9NM09 leishmania	c 751	7	0.5	214	10	Q94H35	Q94H35 oryza sativ
679	7	0.5	206	5	Q9VYC9	Q9VYC9 drosophila	752	7	0.5	215	2	Q9SV19	Q9SV19 arabidopsis
680	7	0.5	206	8	Q8WC92	Q8WC92 leposoma pe	c 753	7	0.5	215	2	Q93A10	Q93A10 thiobacillu
681	7	0.5	207	8	Q9B019	Q9B019 limnodynast	c 754	7	0.5	216	3	Q9USA7	Q9USA7 schizosacch
682	7	0.5	207	8	Q8WCC4	Q8WCC4 arthrosaura	755	7	0.5	216	8	Q9B0L0	Q9B0L0 limnodynast
683	7	0.5	207	8	Q8WCC3	Q8WCC3 gymnoththal	756	7	0.5	216	16	Q98H77	Q98H77 rhizobium l
684	7	0.5	207	8	Q8WCC2	Q8WCC2 colobosaura	757	7	0.5	217	3	Q94290	Q94290 schizosacch
685	7	0.5	207	8	Q8WCC1	Q8WCC1 bachia flav	758	7	0.5	217	8	Q9B0I5	Q9B0I5 limnodynast
686	7	0.5	207	8	Q8WCC0	Q8WCC0 procellosau	759	7	0.5	217	16	Q9RIZ6	Q9RIZ6 streptomyce
687	7	0.5	207	8	Q8WC89	Q8WC89 procellosau	760	7	0.5	218	8	Q9T7Z7	Q9T7Z7 lampropelti
688	7	0.5	207	8	Q8WC88	Q8WC88 psilophthal	761	7	0.5	218	8	Q9T7Z6	Q9T7Z6 lampropelti
689	7	0.5	207	8	Q8WC87	Q8WC87 calyptommat	762	7	0.5	218	8	Q9T7Z5	Q9T7Z5 lampropelti
690	7	0.5	207	8	Q8WC86	Q8WC86 calyptommat	763	7	0.5	218	8	Q9T7Z4	Q9T7Z4 lampropelti
691	7	0.5	207	8	Q8WC85	Q8WC85 micrablepha	764	7	0.5	218	8	Q9T7Z3	Q9T7Z3 lampropelti
692	7	0.5	207	8	Q8WC84	Q8WC84 bachia bres	765	7	0.5	218	8	Q9T7Z2	Q9T7Z2 lampropelti
693	7	0.5	207	8	Q8WC83	Q8WC83 anotosaura	766	7	0.5	218	8	Q9T7Z1	Q9T7Z1 lampropelti
694	7	0.5	207	8	Q8WC82	Q8WC82 neusticurus	767	7	0.5	218	8	Q9T7Z0	Q9T7Z0 lampropelti
695	7	0.5	207	8	Q8WC81	Q8WC81 placosoma c	768	7	0.5	218	8	Q9T7Y9	Q9T7Y9 lampropelti
696	7	0.5	207	8	Q8WC80	Q8WC80 pantodactyl	769	7	0.5	218	8	Q9T7Y8	Q9T7Y8 lampropelti
697	7	0.5	207	8	Q8WC89	Q8WC89 colobodactyl	770	7	0.5	218	8	Q9T7Y7	Q9T7Y7 lampropelti
698	7	0.5	207	8	Q8WC88	Q8WC88 pantodactyl	771	7	0.5	218	8	Q9T7Y6	Q9T7Y6 lampropelti
699	7	0.5	207	8	Q8WC87	Q8WC87 cercosaura	772	7	0.5	218	8	Q9T7Y5	Q9T7Y5 lampropelti
700	7	0.5	207	8	Q8WC86	Q8WC86 pholidobolu	773	7	0.5	218	8	Q9T7Y4	Q9T7Y4 lampropelti
701	7	0.5	207	8	Q8WC85	Q8WC85 heterodactyl	774	7	0.5	218	8	Q9T7Y3	Q9T7Y3 lampropelti
702	7	0.5	207	8	Q8WC84	Q8WC84 colobosau	775	7	0.5	218	8	Q9T7Y2	Q9T7Y2 lampropelti
703	7	0.5	207	8	Q8WC83	Q8WC83 colobosaura	776	7	0.5	218	8	Q9T7Y1	Q9T7Y1 lampropelti
704	7	0.5	207	8	Q8WC82	Q8WC82 tretioscinc	777	7	0.5	218	8	Q9T7Y0	Q9T7Y0 lampropelti
705	7	0.5	207	8	Q8WC81	Q8WC81 iphisa eleg	778	7	0.5	218	8	Q9T7X9	Q9T7X9 lampropelti
706	7	0.5	207	8	Q8WC80	Q8WC80 neusticurus	779	7	0.5	218	8	Q9T7X8	Q9T7X8 lampropelti
707	7	0.5	207	8	Q8WC99	Q8WC99 tretioscinc	780	7	0.5	218	8	Q9T7X7	Q9T7X7 lampropelti
708	7	0.5	207	8	Q8WC98	Q8WC98 bachia dorb	781	7	0.5	218	8	Q9T7X6	Q9T7X6 lampropelti
709	7	0.5	207	8	Q8WC97	Q8WC97 prionodactyl	782	7	0.5	218	8	Q9T7X5	Q9T7X5 lampropelti
710	7	0.5	207	8	Q8WC96	Q8WC96 arthrosaura	783	7	0.5	218	8	Q9B3U2	Q9B3U2 plethodon w
711	7	0.5	207	8	Q8WC95	Q8WC95 prionodactyl	784	7	0.5	218	8	Q9B3T5	Q9B3T5 plethodon s
712	7	0.5	207	8	Q8WC94	Q8WC94 prionodactyl	785	7	0.5	218	8	Q9B0J5	Q9B0J5 limnodynast
713	7	0.5	207	8	Q8WC93	Q8WC93 leposoma os	786	7	0.5	218	8	Q9B3U6	Q9B3U6 plethodon s
714	7	0.5	207	8	Q8WC91	Q8WC91 colobosaura	787	7	0.5	218	8	Q9B3V5	Q9B3V5 aneides lug
715	7	0.5	207	8	Q8WC90	Q8WC90 notobachia	788	7	0.5	218	8	Q9T3U7	Q9T3U7 lampropelti
716	7	0.5	207	8	Q8WC78	Q8WC78 tupinambis	789	7	0.5	218	8	Q9T3U6	Q9T3U6 lampropelti
717	7	0.5	207	8	Q8WC77	Q8WC77 kentropyx c	790	7	0.5	218	8	Q9T350	Q9T350 lampropelti
718	7	0.5	207	8	Q8WC76	Q8WC76 cremidophor	791	7	0.5	218	10	Q93WG8	Q93WG8 oryza sativ
c 719	7	0.5	207	15	Q8UTE4	Q8UTE4 human immun	c 792	7	0.5	218	13	Q9FU72	Q9FU72 cynops pyrr
c 720	7	0.5	207	17	Q8WTP9	Q8WTP9 methanopyru	c 793	7	0.5	218	16	P96227	P96227 mycobacteri
c 721	7	0.5	208	5	Q8WTD3	Q8WTD3 glossina mo	794	7	0.5	219	3	Q12579	Q12579 chaetomium
c 722	7	0.5	208	10	Q9M5W7	Q9M5W7 vaccinium c	795	7	0.5	219	8	Q9M167	Q9M167 agkistrodon
723	7	0.5	208	10	Q04033	Q04033 arabidopsis	796	7	0.5	219	8	Q9B3V4	Q9B3V4 plethodon c
c 724	7	0.5	208	16	Q8XDX2	Q8XDX2 escherichia	797	7	0.5	219	8	Q9B0L1	Q9B0L1 limnodynast
c 725	7	0.5	209	5	Q8SQN4	Q8SQN4 encephalito	798	7	0.5	219	8	Q9B0K9	Q9B0K9 limnodynast
726	7	0.5	209	8	Q9B0J4	Q9B0J4 limnodynast	799	7	0.5	219	8	Q9B0K8	Q9B0K8 limnodynast
727	7	0.5	209	8	Q952E1	Q952E1 lineatriton	800	7	0.5	219	8	Q9B0I6	Q9B0I6 limnodynast
728	7	0.5	209	8	Q8WC79	Q8WC79 pantodactyl	801	7	0.5	219	8	Q9B0I4	Q9B0I4 limnodynast
c 729	7	0.5	209	17	Q27345	Q27345 methanobact	802	7	0.5	219	8	Q9B0I2	Q9B0I2 limnodynast
730	7	0.5	210	5	Q9WZ27	Q9WZ27 drosophila	803	7	0.5	219	8	Q9B3W0	Q9B3W0 aneides fer
731	7	0.5	210	17	Q9HQC4	Q9HQC4 halobacteri	804	7	0.5	219	8	Q9B3W1	Q9B3W1 aneides fer

c 805	7	0.5	220	1	Q9UXQ2	Q9uxq2 methanobact	878	7	0.5	225	5	Q9NG62	Q9ng62 drosophila
c 806	7	0.5	220	5	Q9N7P5	Q9n7p5 leishmania	879	7	0.5	225	8	Q9B0M5	Q9b0m5 limnodynast
c 807	7	0.5	220	8	Q9B4G9	Q9b4g9 eryx miliar	880	7	0.5	225	8	Q9B0M4	Q9b0m4 limnodynast
c 808	7	0.5	220	8	Q9B3V1	Q9b3v1 plethodon h	881	7	0.5	225	8	Q9B0M3	Q9b0m3 limnodynast
c 809	7	0.5	220	8	Q9B309	Q9b3u9 plethodon o	882	7	0.5	225	8	Q9B0M2	Q9b0m2 limnodynast
c 810	7	0.5	220	8	Q9B307	Q9b3u7 plethodon s	883	7	0.5	225	8	Q9B0M1	Q9b0m1 limnodynast
c 811	7	0.5	220	16	Q9CWS8	Q9cms8 pasteurella	884	7	0.5	225	8	Q9B0M0	Q9b0m0 limnodynast
c 812	7	0.5	220	16	Q9K3I0	Q9k3i0 streptomyce	885	7	0.5	225	8	Q9B0L9	Q9b0l9 limnodynast
c 813	7	0.5	220	17	Q9HLA8	Q9hla8 thermoplas	886	7	0.5	225	8	Q9B0L8	Q9b0l8 limnodynast
c 814	7	0.5	220	17	Q8U207	Q8u207 pyrococcus	887	7	0.5	225	8	Q9B0L7	Q9b0l7 limnodynast
c 815	7	0.5	221	8	Q9MI68	Q9mi68 agkistrodon	888	7	0.5	225	8	Q9B0L6	Q9b0l6 limnodynast
c 816	7	0.5	221	8	Q9B3T0	Q9b3t0 plethodon v	889	7	0.5	225	8	Q9B0L5	Q9b0l5 limnodynast
c 817	7	0.5	221	8	Q9B3V9	Q9b3v9 aneides fla	890	7	0.5	225	8	Q9B3T8	Q9b3t8 plethodon e
c 818	7	0.5	221	12	Q8S299	Q8s299 orf virus	891	7	0.5	225	8	Q9B3T7	Q9b3t7 plethodon e
c 819	7	0.5	221	16	Q985A6	Q985a6 rhizobium l	892	7	0.5	225	8	Q9B3T6	Q9b3t6 plethodon e
c 820	7	0.5	221	16	Q8YKM9	Q8ykm9 anabaena sp	893	7	0.5	225	8	Q9B3T2	Q9b3t2 plethodon n
c 821	7	0.5	221	16	Q9S285	Q9s285 streptomyce	894	7	0.5	225	8	Q9B3T1	Q9b3t1 plethodon n
c 822	7	0.5	222	2	Q9L7R0	Q9l7r0 streptococc	895	7	0.5	225	8	Q9B0L4	Q9b0l4 limnodynast
c 823	7	0.5	222	2	Q9X622	Q9x622 bacillus sp	896	7	0.5	225	8	Q9B0K7	Q9b0k7 limnodynast
c 824	7	0.5	222	5	Q9NLE8	Q9nle8 leishmania	897	7	0.5	225	8	Q9B0K6	Q9b0k6 limnodynast
c 825	7	0.5	222	8	Q9B3U5	Q9b3u5 plethodon i	898	7	0.5	225	8	Q9B0K5	Q9b0k5 limnodynast
c 826	7	0.5	222	8	Q95SR6	Q95sr6 micrurus co	899	7	0.5	225	8	Q9B0K4	Q9b0k4 limnodynast
c 827	7	0.5	222	8	Q95SR5	Q95sr5 micrurus fr	900	7	0.5	225	8	Q9B0K2	Q9b0k2 limnodynast
c 828	7	0.5	222	8	Q95SR4	Q95sr4 micrurus fr	901	7	0.5	225	8	Q9B0K1	Q9b0k1 limnodynast
c 829	7	0.5	222	8	Q95SR3	Q95sr3 micrurus al	902	7	0.5	225	8	Q9B0J9	Q9b0j9 limnodynast
c 830	7	0.5	222	8	Q95SR2	Q95sr2 micrurus al	903	7	0.5	225	8	Q9B0J8	Q9b0j8 limnodynast
c 831	7	0.5	222	8	Q95SR1	Q95sr1 micrurus al	904	7	0.5	225	8	Q9B0J6	Q9b0j6 limnodynast
c 832	7	0.5	222	8	Q95SR0	Q95sr0 micrurus al	905	7	0.5	225	8	Q9B0J3	Q9b0j3 limnodynast
c 833	7	0.5	222	8	Q95S09	Q95sq9 micrurus ba	906	7	0.5	225	8	Q9B0J2	Q9b0j2 limnodynast
c 834	7	0.5	222	8	Q95S08	Q95sq8 micrurus py	907	7	0.5	225	8	Q9B0J1	Q9b0j1 limnodynast
c 835	7	0.5	222	8	Q95S07	Q95sq7 micrurus le	908	7	0.5	225	8	Q9B0J0	Q9b0j0 limnodynast
c 836	7	0.5	222	8	Q95S06	Q95sq6 micrurus le	909	7	0.5	225	8	Q9B0I8	Q9b0i8 limnodynast
c 837	7	0.5	222	8	Q95S05	Q95sq5 micrurus le	910	7	0.5	225	8	Q9B0I7	Q9b0i7 limnodynast
c 838	7	0.5	222	8	Q95S04	Q95sq4 micrurus le	911	7	0.5	225	8	Q9B0I3	Q9b0i3 limnodynast
c 839	7	0.5	222	8	Q95S03	Q95sq3 micrurus le	912	7	0.5	225	8	Q9B0I1	Q9b0i1 limnodynast
c 840	7	0.5	222	8	Q95S02	Q95sq2 micrurus ib	913	7	0.5	225	8	Q9B3V6	Q9b3v6 aneides lug
c 841	7	0.5	222	8	Q95S01	Q95sq1 micrurus de	914	7	0.5	225	8	Q9B3V8	Q9b3v8 aneides fla
c 842	7	0.5	222	8	Q95S00	Q95sq0 micrurus he	915	7	0.5	225	12	Q9PWZ5	Q9pwz5 shope fibro
c 843	7	0.5	222	8	Q95S09	Q95sq9 micrurus sp	c 916	7	0.5	226	8	Q953I5	Q953i5 echinosorex
c 844	7	0.5	222	8	Q95SP8	Q95sp8 micrurus su	917	7	0.5	226	8	Q9B3W3	Q9b3w3 ensatina es
c 845	7	0.5	222	8	Q94PE7	Q94pg7 micrurus br	918	7	0.5	226	8	Q9B3W4	Q9b3w4 ensatina es
c 846	7	0.5	222	16	Q9HWM7	Q9hwm7 pseudomonas	919	7	0.5	226	8	Q8SJ12	Q8sj12 ursus arcto
c 847	7	0.5	223	4	Q9H687	Q9h687 homo sapien	c 920	7	0.5	226	8	Q8SJH0	Q8sjh0 ursus marit
c 848	7	0.5	223	8	Q9B0L3	Q9b0l3 limnodynast	c 921	7	0.5	226	16	Q9I103	Q9i103 pseudomonas
c 849	7	0.5	223	8	Q94X31	Q94y31 crotalus vi	922	7	0.5	227	3	Q8X0K5	Q8x0k5 neurospora
c 850	7	0.5	223	8	Q94Y30	Q94y30 crotalus vi	923	7	0.5	227	5	Q9BL42	Q9bl42 caenorhabdi
c 851	7	0.5	223	8	Q94Y29	Q94y29 crotalus vi	c 924	7	0.5	227	5	Q9VGF4	Q9vgf4 drosophila
c 852	7	0.5	223	8	Q94Y28	Q94y28 crotalus vi	c 925	7	0.5	227	8	O21346	O21346 euhadra her
c 853	7	0.5	223	8	Q94Y27	Q94y27 crotalus vi	926	7	0.5	227	8	Q9B3V7	Q9b3v7 aneides har
c 854	7	0.5	223	8	Q94Y26	Q94y26 crotalus vi	c 927	7	0.5	227	11	Q9CYL2	Q9cyl2 mus musculu
c 855	7	0.5	223	8	Q94Y25	Q94y25 crotalus vi	c 928	7	0.5	227	11	Q9CPQ0	Q9cpq0 mus musculu
c 856	7	0.5	223	8	Q94Y24	Q94y24 crotalus vi	c 929	7	0.5	227	16	Q9CD41	Q9cd41 mycobacteri
c 857	7	0.5	223	8	Q94Y23	Q94y23 crotalus vi	c 930	7	0.5	227	16	Q8XIA0	Q8xia0 clostridium
c 858	7	0.5	223	8	Q94Y22	Q94y22 crotalus vi	931	7	0.5	228	3	O74463	O74463 schizosacch
c 859	7	0.5	223	8	Q94Y21	Q94y21 crotalus vi	932	7	0.5	228	8	Q9T7X4	Q9t7x4 coluber con
c 860	7	0.5	223	8	Q94Y20	Q94y20 crotalus vi	933	7	0.5	228	8	Q9B3W5	Q9b3w5 ensatina es
c 861	7	0.5	223	8	Q94Y19	Q94y19 crotalus vi	934	7	0.5	228	10	Q94H60	Q94h60 oryza sativ
c 862	7	0.5	223	8	Q94Y18	Q94y18 crotalus vi	c 935	7	0.5	228	11	Q9CYL8	Q9cyl8 mus musculu
c 863	7	0.5	223	8	Q94Y17	Q94y17 crotalus vi	c 936	7	0.5	228	11	Q9CPQ2	Q9cpq2 mus musculu
c 864	7	0.5	223	8	Q94Y16	Q94y16 crotalus vi	c 937	7	0.5	228	11	Q920H9	Q920h9 rattus norv
c 865	7	0.5	223	8	Q94Y15	Q94y15 crotalus vi	938	7	0.5	228	16	Q9RW00	Q9rw00 deinococcus
c 866	7	0.5	223	8	Q94Y14	Q94y14 crotalus vi	939	7	0.5	228	16	Q9CDY8	Q9cdy8 lactococcus
c 867	7	0.5	223	8	Q94Y13	Q94y13 crotalus vi	940	7	0.5	229	8	Q952E6	Q952e6 lineatriton
c 868	7	0.5	223	8	Q94Y12	Q94y12 crotalus sc	941	7	0.5	229	8	Q8WA06	Q8wa06 dendroaspis
c 869	7	0.5	223	8	Q94Y11	Q94y11 crotalus du	942	7	0.5	229	8	Q8W9Z7	Q8w9z7 ophiophagus
c 870	7	0.5	223	10	Q9W5C3	Q9w5c3 solanum cha	c 943	7	0.5	229	10	Q9LFG5	Q9lpg5 arabiopsis
c 871	7	0.5	223	11	Q63211	Q63211 rattus norv	c 944	7	0.5	229	16	O8YLP4	O8ylp4 anabaena sp
c 872	7	0.5	224	8	Q9MLQ4	Q9mlq4 limulus pol	c 945	7	0.5	229	17	Q8ZW30	Q8zw30 pyrobaculum
c 873	7	0.5	224	8	Q9B3T9	Q9b3t9 plethodon e	c 946	7	0.5	230	5	Q9TZ76	Q9tz76 caenorhabdi
c 874	7	0.5	224	8	Q9B0L2	Q9b0l2 limnodynast	947	7	0.5	230	8	Q9MEB5	Q9meb5 agkistrodon
c 875	7	0.5	224	8	Q9B3W6	Q9b3w6 ensatina es	948	7	0.5	230	8	Q9MI71	Q9mi71 agkistrodon
c 876	7	0.5	224	8	Q8W9Z3	Q8w9z3 walterinnes	949	7	0.5	230	8	Q9MI66	Q9mi66 agkistrodon
c 877	7	0.5	224	16	Q92JH1	Q92jh1 rickettsia	950	7	0.5	230	8	Q9MI76	Q9mi76 crotalus ti

951	7	0.5	230	8	Q9M175	Q9m175
952	7	0.5	230	8	Q9M174	Q9m174
953	7	0.5	230	8	Q9M173	Q9m173
954	7	0.5	230	8	Q9M170	Q9m170
955	7	0.5	230	8	Q9M169	Q9m169
956	7	0.5	230	8	Q9M163	Q9m163
957	7	0.5	230	8	Q94UE0	Q94ue0
958	7	0.5	230	8	Q8W925	Q8w925
959	7	0.5	230	8	Q8SKS0	Q8skso
960	7	0.5	230	8	Q8SKR7	Q8skr7
961	7	0.5	231	8	Q9MDM6	Q9mdm6
962	7	0.5	231	8	Q79627	Q79627
963	7	0.5	231	8	Q79628	Q79628
964	7	0.5	231	8	Q79629	Q79629
965	7	0.5	231	8	Q79630	Q79630
966	7	0.5	231	8	Q79631	Q79631
967	7	0.5	231	8	Q79632	Q79632
968	7	0.5	231	8	Q79633	Q79633
969	7	0.5	231	8	Q79634	Q79634
970	7	0.5	231	8	Q79472	Q79472
971	7	0.5	231	8	Q79474	Q79474
972	7	0.5	231	8	Q79475	Q79475
973	7	0.5	231	8	Q79476	Q79476
974	7	0.5	231	8	Q79477	Q79477
975	7	0.5	231	8	Q79478	Q79478
976	7	0.5	231	8	Q79479	Q79479
977	7	0.5	231	8	Q79480	Q79480
978	7	0.5	231	8	Q79482	Q79482
979	7	0.5	231	8	Q79483	Q79483
980	7	0.5	231	8	Q79484	Q79484
981	7	0.5	231	8	Q79485	Q79485
982	7	0.5	231	8	Q79487	Q79487
983	7	0.5	231	8	Q79489	Q79489
984	7	0.5	231	8	Q79491	Q79491
985	7	0.5	231	8	Q79492	Q79492
986	7	0.5	231	8	Q79493	Q79493
987	7	0.5	231	8	Q79494	Q79494
988	7	0.5	231	8	Q79496	Q79496
989	7	0.5	231	8	Q79497	Q79497
990	7	0.5	231	8	Q79498	Q79498
991	7	0.5	231	8	Q79500	Q79500
992	7	0.5	231	8	Q79501	Q79501
993	7	0.5	231	8	Q79502	Q79502
994	7	0.5	231	8	Q79503	Q79503
995	7	0.5	231	8	Q9T621	Q9t621
996	7	0.5	231	8	Q9T620	Q9t620
997	7	0.5	231	8	Q9T619	Q9t619
998	7	0.5	231	8	Q9T618	Q9t618
999	7	0.5	231	8	Q9T617	Q9t617
1000	7	0.5	231	8	Q9T616	Q9t616
					Q9T7X3	Q9t7x3

ALIGNMENTS

RESULT 1
Q96KF5 PRELIMINARY; PRT; 1320 AA.
AC Q96KF5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Myopalladin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RX MEDLINE=21206024; PubMed=11309420;
RA Bang M.L., Mudry R.E., McElhinny A.S., Trombitas K., Geach A.J.,
RA Yamasaki R., Sorimachi H., Granzier H., Gregorio C.C., Labeit S.;
RT "Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple

RT roles in 2-disc and I-band protein assemblies.";
RL J. Cell Biol. 153:413-428(2001).
DR EMBL; AF328296; AAK50625.1; -.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 5.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
SQ SEQUENCE 1320 AA; 145379 MW; A6579FB164D33B6E CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1320
Score: 770.00 Matches: 1314
Percent Similarity: 99.24% Conservative: 0
Best Local Similarity: 99.24% Mismatches: 6
Query Match: 58.51% Indels: 8
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q96KF5 (1-1320)

Qy	1	ATGCAAGACGACGACATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT	60
Db	1	MetGlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr	20
Qy	61	TTAGCTGAACACGACATCGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCTCAAC	120
Db	21	LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn	40
Qy	121	CCTTGCCATTTCGGCAGCTCTCTGGGGCGCTCAAGGAGCGGCGGCAAGATGACCTT	180
Db	41	ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu	60
Qy	181	CCAGATCTTTTCAGCTTCTGAGCCAAAGAAATTAGACGAAAGTGTCAATTTGGCAAGA	240
Db	61	ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg	80
Qy	241	CTGCCATCAATTACGACCCTTTGGAGAGCGAGATGAAACTCAAGCTAGAAAACGACTT	300
Db	81	LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu	100
Qy	301	TCCTCTCATCAGTGAACACTCACCTAATTAGTTTTCAGCCTAACTTCTGCCAGAT	360
Db	101	SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp	120
Qy	361	AACCTCTGAAGTCCACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT	420
Db	121	AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr	140
Qy	421	TGTTCTGAACCCAGTCCAAAAGTATTTTAAATTAAGCTGCCGACTTCATTGAAGAG	480
Db	141	CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu	160
Qy	481	CTATCTCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCGCTGCAAAAAC	540
Db	161	LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn	180
Qy	541	CACAAGAGTAACTGGAATCTCAAAACAAAAGTTATGTCAGGAAACACAGCTCCAGCTTTC	600
Db	181	HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer	200
Qy	601	GATCTGTGAGAAGACAGAGAAAGATCTTCTGTTCCCATCCCTATCCCTGCGGATAC	660
Db	201	AspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThrArg	220
Qy	661	GATTAATCAAGTGAATCAGCCCTCGGAACACAGCAGCAAGAGCGCTGAAGCGGACAG	720
Db	221	AspAsnGluValAsnHisAlaLeuGluGlnGlnGlnAlaLysArgArgGluAlaGluGln	240
Qy	721	GCTGCCAGTGGCGGCTGGTGGAGACACTACACAGGGCTTTCCCTTCATCTCTGTAC	780
Db	241	AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr	260
Qy	781	TATGAAGAAGCTCTGGGGCAACCTCCCGGTTCTACTCAAAAGTTACGGACGAGAGATT	840

Db 261 TyrGluGluProLeuGluGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
QY 841 CCAGAAGGAACTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACACCTCAAGTA 900
Db 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
QY 901 AGGTGGTACTGTGAAGCGAAGGAGCTTGAAATTCGCCAGATATTACATCGTCCAGGCA 960
Db 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
QY 961 GGAATCTGCATCTACCTGACCATTGGGAAGCCTTTGAGAGGACACAGACGCTATTCC 1020
Db 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
QY 1021 TGCCTTTGCTTCTTAACATCATGGCAGAGATTCGACTTCTGCTCAGATTTATATAGAAGG 1080
Db 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
QY 1081 GTTCTTCTTCTGACTCAGAAGCGACCCCTTAACAAGGAAGAGATGAATCGAATCCAGAAG 1140
Db 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
QY 1141 CCAATAGAGTGCATCTCTCCATCTACTCTGCGAGTCAATTCCTCAGCAGTACCCCAA 1200
Db 381 ProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValProGln 400
QY 1201 GCCCAGCATTTGCTGGCCCAACCTCTGTGGCAACCATCCAGCAGTGTAGAGCCCCACC 1260
Db 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
QY 1261 AATTACTTGCAGGGATTGGATGGAACCTATCATTTGCAGCTCCTGTGTTTACAAAGATG 1320
Db 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaProValPheThrLysMet 440
QY 1321 CTCAAAAATTGTCAGCTTCTGAGGTCAGCTGCTGTTGCTTTGAATGCAGAGTAAAGGA 1380
Db 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
QY 1381 GCTCCATCTCTTAAGTTGAGTGGTATAGAGAAGGACTTTAATAGAAGATTCTCCAGAT 1440
Db 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
QY 1441 TTTAGGATTTACAGAAAAACCTCGATCCATGGCAGAGCAGAGGAGATTTGCACCTTG 1500
Db 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu 500
QY 1501 GTCATTGCTCAGGTGTTGCAGAGATTCTGGGTGCTTACATGTACTGCAACACACAAA 1560
Db 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
QY 1561 TACGGCAGCTGTCAAGCATTCACAGCTGCACGTGCAGAGGAAATGAGGACCTCAGCAAC 1620
Db 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
QY 1621 AACGGGTCTCTTCACTCAGCAACTCYACCACCAACCTGGCAGCTAFTTAGCCACACAGCCC 1680
Db 541 AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnPro 560
QY 1681 TCCCCACCCACATCAGAGCCTCATCTGTGGAAACACACCCCAACCCCAACCTCAGAGGG 1740
Db 561 SerProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGly 580
QY 1741 GTTCTGGTGAACACATGAGCCCGTCCAGCTCCAGGATTTGGCTTCTGTCACCTTC 1800
Db 581 ValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHisPhe 600
QY 1801 AACCTGCTGAAGATGACAAAAGGAAGTGAAGCATCTCCAGGCTGGTGGTGAGCACCC 1860
Db 601 AsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThr 620
QY 1861 AGACAGACCAAGCCCGATTCTTTSCAGGAGAGGTTCAACGGCAGACGACACAAAAACCCC 1919
Db 621 ArgGlnThrArgProAspSerPhe-GlnGluArgPheAsnGlyGlnAlaThrLysThrPr 640

QY 1920 AGAGCCTTCTTTTCCCC - GTGAAAGAGCCCTCCAGTCTTGCCAAACCCAAACTTGATT 1978
Db 640 OGUProSer - SerProValLysGluProProValLeuAlaLysProLysLeuAspS 660
QY 1979 CCACTCAGTTACACAGCTTCATAACCAAGTCTTACTGGAAACCAACCAATTCAAAACC 2038
Db 660 erThrGlnLeuGlnLeuHisAsnGlnValLeuLeuGluGlnHisGlnLeuGlnAsn 680
QY 2039 CACCTCTTCATCTCCTAAGGAGTTTCCITTCARCATGACTGTTTGAACCTCAATGCTC 2098
Db 680 roProProSerSerProLysGluPheProPheAsnMetThrValLeuAsnSerAsnAla 700
QY 2099 CCCACGGGTGACACATCCARTAAGCAGGTGAAGGCTCTCTTCATCACACGCTTCAGCT 2158
Db 700 roProAlaValThrThrSerAsnLysGlnValLysAlaProSerSerGlnThrPheSerL 720
QY 2159 TGCCCGCGCCGAAGTATTTTCCCTCCACGAACACCCGCGAGCAACTGTGGCCCCCT 2218
Db 720 euAlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProS 740
QY 2219 CCAGCTCTCCGGTGTTCACCTTTCAGCAGCAGCTCCTCAAATATTCAGAGACAGTGAGCA 2278
Db 740 erSerSerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerL 760
QY 2279 AAAAAAGCCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAATCTCCAGGAGGCTTCCA 2338
Db 760 ysluSerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerI 780
QY 2339 TCCAAATAGCCACTCCCAACAGGCGCCAAACAGAACCCACACCACTTTCACATTTT 2398
Db 780 leGlnAsnGluProLeuProProGlyProThrGluProThrProProPheThrPheS 800
QY 2399 CCATCCCCAGC - GGAACACAGTTTCAGCCCGCTGTGTGTCGCCAATTCCTGTCTCCT 2457
Db 800 erIle - ProArgGlyAsnGlnPheGlnProArgCysValSerProIleProValSerPro 819
QY 2458 ACCAGCGGATTCAGAACCCAGTGGCTTCTCAGCTCTGTTCTGCTTCTCCTCCCTGCC 2517
Db 820 ThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAla 839
QY 2518 ATCCCAACCCACAAATGCCAT - GGGGTGCTAGAAAGTGCACCATCCATGCCATCCCAAGG 2576
Db 840 ileProProThrAsnAlaMetGly - LeuProArgSerAlaProSerMetProSerGlnG 859
QY 2577 ATTAGCGAAGAAAAATACAAAGTCTCCTCAACAGTGAATGATCATACATTCTGTGAAC 2636
Db 859 yLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAsnIleArgGluTh 879
QY 2637 TAAGAACGCGAGTGCAGACTTGGGAAAAAATAACTTTTCAGTCATGTCAGACCCAAA 2696
Db 879 rLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAs 899
QY 2697 CCAGCAGGAGTACAAAATTTCAAGCTTTGACGAGAGGCTGATGAATGAATAGAGTTTCG 2756
Db 899 nGlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetasngluIleGluPheAr 919
QY 2757 CTTTGAACGTACTCTCTGTTGATGAATCAGATGATGAATTCACATCATGATGAGATCCCC 2816
Db 919 gLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProTh 939
QY 2817 GGCACAGTGTATTCCTCCCATCTTTGACAAGAGACTCAAGCACTTCCTCGGCTCACAGAGG 2876
Db 939 rGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluG 959
QY 2877 CTCTCAGTCACATTCACCTGCAAAATTTGTTGGGATACCTGTTCCAAAGGTTTACTGGT 2936
Db 959 ySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPh 979
QY 2937 CAAAGATGGGAAGCAGATTTCTAAGAAATGACACTGCAAAATGAGGCGAGAGGAGA 2996
Db 979 eLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluGluLys 999

```
QY 2997 TGGGACATGCTCTCTGCACATTTGAATCCATACACATGATGACGATGGCAGCACTACACCAT 3056
DB 999 pGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIle 1019
QY 3057 CATGGCAGCCAAACCCAGGGGAGAATCAGCTGTCTGGCCACTTGATGGTACAAAGTTT 3116
DB 1019 eMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLe 1039
QY 3117 GCCCATTCGCGAGTCGGGTAACTCTGCTGGTCACTCAGAGGGAAGATCCCGAGTGCA 3176
DB 1039 uProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgIleArgSerArgValGI 1059
QY 3177 AGAAGAGACAAAGAGCCCTACAGSACGCTTTTCCGACACATTTCTCGCAGGCTCC 3236
DB 1059 nGluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPr 1079
QY 3237 TGGGGATATGGTACGTATGAGGGGCGCCCTCTGCTGGCTGGACTGTAAGGTGAGTGT 3296
DB 1079 oGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLe 1099
QY 3297 ACCGCCCCGGGAGCTGACATGGCTACTCAATGGCCAACTGTGTACACAGATGCTCCCA 3356
DB 1099 uProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHi 1119
QY 3357 CAAGATGCTGTGAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCACTCACTCAGCG 3416
DB 1119 sLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnAr 1139
QY 3417 CGACGCGAGGACCTTAAAGTCATCGCTACCAACAAACCGGCGAGAAATCTTTAGTCT 3476
DB 1139 gAspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLe 1159
QY 3477 GGAGCTCTCTGTAGTAGCAAAAGAGGTGAAGAAAGCACTGTGATCTCTGGAGAACTACA 3536
DB 1159 uGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGI 1179
QY 3537 GAATCGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGCGGTAGGCAATGCC 3596
DB 1179 nAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPr 1199
QY 3597 CCCACCTGTGTTCTTACCTGGAAGAAGACATGAGACCATCCCTTGCACCGAGAGAGAT 3656
DB 1199 oProProValPheTrpTrpLysLysAspAsnGluThrIleProCysThrArgGluArgII 1219
QY 3657 CAGTATGACACGACGACACACAGGTTGCTGCTGCTCTCTATTGACGACGACCAAGAAATC 3716
DB 1219 eSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSe 1239
QY 3717 AGACGCTGGATGTACAGTTGTGACGCCAAGATGAAGCGCGCATCGTGTGCTGCACTGC 3776
DB 1239 rAspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAl 1259
QY 3777 CAGGCTGSATATATACGCTCAGTGGACCATCAGATCCACCGCCCATGTCTGTCGCGCC 3836
DB 1259 aArgLeuAspIleTrpAlaGlnTrpHisGlnIleProProMetSerValArgPr 1279
QY 3837 CAGTGGCAGTCTGCTACGGATCTCTCACCAGTAAAGGACTTGACATATTTCTGCTTTTC 3896
DB 1279 oSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSe 1299
QY 3897 CTCCATGGAAGACGATGGTGTATTCATGCTTCTCGGAGTGCTAGTGAGAGTGATGA 3956
DB 1299 rSerMetGluSerThrMetValTyrSerCysSerArgSerValValGluSerAspGI 1319
QY 3957 ACTT 3960
DB 1319 uLeu 1320
RESULT 2
Q96K90 PRELIMINARY; PRT; 507 AA.
ID Q96K90
AC Q96K90;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CNDA FLJ14437 fis, clone HEMBB100083, weakly similar to myosin light
DE chain kinase, smooth muscle and NON-muscle isozymes
DE (EC 2.7.1.117).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027343; BAB55048.1;
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 2.
SQ SEQUENCE 507 AA; 56185 MW; BD484878845DF39 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 507
Score: 494.00 Matches: 494
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.54% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q96K90 (1-507)

QY 1 ATCAAGACGACGACGATAGAGCTTCTTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 60
DB 1 MetGlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
QY 61 TTAGCTAAACCCAGACATCGGGGAAACAATGAGAGGAGTCGAGGGAGCCCTCTCTCAAC 120
DB 21 LeuAlaGluThrArgHisArgGlyAsnGluArgSerArgAlaGluProSerSerAsn 40
QY 121 CCTTGCCATTTCGGCAGCTCTCTTCTGGGCGCGCTGAAGGAGGCGGAGGCAAGATGACCTT 180
DB 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu 60
QY 181 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAGATTAGACGAAAGTGTCAATTTGGCAAGA 240
DB 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
QY 241 CTGGCCATCAATTACGACCCCTTTGGAGAAGCGCAGATCAAACTCAAGCTAGAAAACGACTT 300
DB 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
QY 301 TCTCCTCATCAGATGAACACACTCACCTTAATTTAAGTTTTCAGCCTTAATCTCTGCCAGAT 360
DB 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
QY 361 AACCTCGAAGTCCCACTCCAGCTCTTAAAGAAAGCCCGAGGAGGCAAAAGCCACAGTAT 420
DB 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
QY 421 TGTCTGAAACCCAGTCCAAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGAGAG 480
DB 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
QY 481 CTATCTCCCTCTTCAAAATCCACAGCTCCAAAAGGATTAGACTCGTCTGCCTGCTCAAAAC 540
DB 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
QY 541 CACAAGAGTAACTGGGAATCTCAAAACAAAGATTATGACGAGAAACAGCTCCAGCTTCTCA 600
```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Liu T., Zhang J., Ye M., Zhang Q., Fu G., Zhou J., Wu J., Shen Y.,
RA Yu M., Chen S., Mao M., Chen Z.;
RA "Human SIH002 gene";
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RL
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077041; AAD27774.1; -;
DR EMBL; BC013867; AAH13867.1; -;
DR HSSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IGC2; 3.
KW Immunoglobulin domain.
SQ SEQUENCE 385 AA; 42932 MW; 15853FB78B8E41B9 CRC64;

Alignment Scores:
Pred. No.: 1.76e-10 Length: 385
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9UQF5 (1-385)

QY 3034 GATGACGATGCGCAACTACACCATCATGCGAGCAACCCAGGAGATCAGCTGT 3030
Db 70 AspaspaspGlyasnrythrileMetAlaAasProGlnGlyArgIleSerCys 88

RESULT 4
Q9Y3E9 PRELIMINARY; PRT; 404 AA.
ID Q9Y3E9 PRELIMINARY;
AC Q9Y3E9; O9UGAO;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CGI-151 protein.
GN DKFZP586L0518.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=2072150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of Novel Human Genes Evolutionarily Conserved in
RT Caenorhabditis elegans by Comparative Proteomics.";
RL Genome Res. 10:703-713(2000).
RN [2]
RN SEQUENCE OF 260-404 FROM N.A.
RC TISSUE=UTERUS;
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151909; AAD34146.1; -;
DR EMBL; AL050093; CAB43265.1; -;
DR HSSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IGC2; 3.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 404 AA; 45005 MW; 729696C78FFC2B5F CRC64;

Alignment Scores:

Pred. No.: 1.75e-10 Length: 404
 Score: 19.00 Matches: 19
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9Y3E9 (1-404)

QY 3034 GATGACGATGCACTACACCATCATGCGACCCAGGGGAGAAATCAGCTGT 3090
 Db 89 AspAspAspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 107

RESULT 5

Q9Y2J6 ID Q9Y2J6 PRELIMINARY; PRT; 772 AA.
 AC Q9Y2J6
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE KIAA0992 protein (Fragment).
 GN KIAA0992
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;

RN [1]

RC SEQUENCE FROM N.A.

RT TISSUE=BRAIN;

RX MEDLINE=99246063; PubMed=10231032;

RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,

RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIII.

RT The complete sequences of 100 new cDNA clones from brain which code

RL DNA Res. 6:63-70(1999).

DR EMBL; AB023209; BAA76836.1; -

DR HSP; P56276; ITLK.

DR InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR002965; P_rich_extensn.

DR Pfam; PF00047; Ig; 3.

DR PRINTS; PR01217; PRICHEXTENS.

DR SMART; SM00408; IGC2; 3.

KW Immunoglobulin domain.

FT NON_TER 1

SQ SEQUENCE 772 AA; 83643 MW; EE5EE2BC79C7492B CRC64;

Alignment Scores:

Pred. No.: 1.6e-10 Length: 772
 Score: 19.00 Matches: 19
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9Y2J6 (1-772)

QY 3034 GATGACGATGCACTACACCATCATGCGACCCAGGGGAGAAATCAGCTGT 3090
 Db 457 AspAspAspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 475

RESULT 6

Q8WX93 ID Q8WX93 PRELIMINARY; PRT; 1106 AA.
 AC Q8WX93
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DE Myoneurin.
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 GN MYN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lockwood S.K., Sims K.B.;

RT "A protein with two immunoglobulin-like domains interacts with the

RL Norrie disease gene product norrin.";

RT Am. J. Hum. Genet. 61:A158-A158(1997).

RN [2]

RP SEQUENCE FROM N.A.

RA Lockwood S.K.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF464873; AAL69964.1; -

DR InterPro; IPR003599; Ig_c2.

DR InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR000634; S/T_dehydrtse.

DR Pfam; PF00047; Ig; 5.

DR SMART; SM00409; IGC2; 5.

DR SMART; SM00408; IGC2; 5.

DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_1.

SQ SEQUENCE 1106 AA; 122019 MW; 80B52268DD90E7C6 CRC64;

Alignment Scores:

Pred. No.: 1.52e-10 Length: 1106
 Score: 19.00 Matches: 19
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q8WX93 (1-1106)

QY 3034 GATGACGATGCACTACACCATCATGCGACCCAGGGGAGAAATCAGCTGT 3090
 Db 844 AspAspAspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 862

RESULT 7

Q9CWW1 ID Q9CWW1 PRELIMINARY; PRT; 379 AA.
 AC Q9CWW1
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 2410003B16Rik protein.
 GN 2410003B16Rik.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;

RC MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito K.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).
 DR EMBL; AK010350; BAB26871.1; -.
 DR HSSP; P56276; 1TLK.
 DR MGD; MGI:1919583; 2410003B16Rik.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003600; Ig_Like.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00409; Ig; 3.
 DR SMART; SM00408; IGC2; 3.
 DR SMART; SM00410; IG_Like; 1.
 KW Immunoglobulin domain.
 SQ SEQUENCE 379 AA; 42157 MW; 42A96BE0AAC8BF25 CRC64;

Alignment Scores:

Pred. No.: 2,95e-07 Length: 379
 Score: 16.00 Matches: 16
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.22% Indels: 0
 DB: 11 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9CWL1 (1-379)

QY 3034 GATGACGATGGCAACTACACATCATCGGACCAACCCCGAGGGAGA 3081
 |||||
 Db 64 AspAspGlyAsnThrIleMetAlaAlaAsnProGlnGlyarg 79

RESULT 8

Q9ET54 PRELIMINARY; PRT; 492 AA.
 AC Q9ET54;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Actin-associated protein palladin (Fragment).
 GN 2410003B16Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.

RC STRAIN=SWISS WEBSTER/NIH;
 RX MEDLINE=20391984; PubMed=10931874;
 RA Parast M.M., Okey C.A.;
 RT "Characterization of palladin, a novel protein localized to stress
 RT fibers and cell adhesions.";
 RL J. Cell Biol. 150:643-656(2000).
 DR EMBL; AF205078; AAG00078.1; -.
 DR HSSP; P56276; 1TLK.
 DR MGD; MGI:1919583; 2410003B16Rik.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003600; Ig_Like.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00409; Ig; 3.
 DR SMART; SM00408; IGC2; 3.
 DR SMART; SM00410; IG_Like; 1.
 KW Immunoglobulin domain.
 FT NON_TER 1
 SQ SEQUENCE 492 AA; 55072 MW; BC59E5B3E3BAEBD5 CRC64;

Alignment Scores:

Pred. No.: 2,85e-07 Length: 492
 Score: 16.00 Matches: 16
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.22% Indels: 0
 DB: 11 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9ET54 (1-492)

QY 3034 GATGACGATGGCAACTACACATCATCGGACCAACCCCGAGGGAGA 3081
 |||||
 Db 177 AspAspGlyAsnThrIleMetAlaAlaAsnProGlnGlyarg 192

RESULT 9

Q9YG65 PRELIMINARY; PRT; 156 AA.
 ID Q9YG65;
 AC Q9YG65;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE0036.
 GN APE0036.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000058; BAA78945.1; -.
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 156 AA; 16252 MW; C0733DF770584F53 CRC64;

Alignment Scores:

Pred. No.: 11.1 Length: 156
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.70% Indels: 0
 DB: 17 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9YG65 (1-156)

QY 752 GTAGTGTCTCCACGACCGCTCACTG 726
 |||||
 Db 128 ValValSerProAlaAlaSerLeu 136

RESULT 10

Q8TXM0 PRELIMINARY; PRT; 239 AA.
 ID Q8TXM0;
 AC Q8TXM0;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Predicted phosphate-binding enzyme of the TIM-barrel fold.
 GN MK0642.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyraceae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natile D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL; AE010357; AA001857.1; -.
 KW Complete proteome.
 SQ SEQUENCE 239 AA; 25052 MW; 179093DD7872137 CRC64;

Alignment Scores:

Pred. No.: 10.4 Length: 239
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.70% Indels: 0
 DB: 17 Gaps: 0

US-09-818-990B-1 (1-3963) x Q8TXM0 (1-239)

QY 2865 CCGGAAGTGTGAGTCTCTGTGCAAA 2839

Db 29 ProgluValLeuGluSerLeuVallys 37

RESULT 11

Q9EV74 ID Q9EV74 PRELIMINARY; PRT; 243 AA.
 AC Q9EV74;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BphS protein.
 DE BphS
 GN BPHS
 OS Alkaligenes eutrophus (Ralstonia eutropha).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A5; TRANSPOSON-TN4371;
 RC MEDLINE=97188526; PubMed=9037111;
 RA Merlin C., Springael D., Mergey M., Toussaint A.;
 RA "Organisation of the bph gene cluster of transposon Tn4371, encoding
 RT enzymes for the degradation of biphenyl and 4-chlorobiphenyl
 RT compounds";
 RL Mol. Gen. Genet. 253:499-506(1997).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=A5; TRANSPOSON-TN4371;
 RC MEDLINE=20092484; PubMed=10628862;
 RA Mouz S., Merlin C., Springael D., Toussaint A.;
 RA "A GntR-like negative regulator of the biphenyl degradation genes of
 RT the transposon Tn4371";
 RL Mol. Gen. Genet. 262:790-799(1999).
 DR EMBL; AJ012075; CAC05302.1; -.
 DR InterPro; IPR000524; HTH_GNTR.
 DR SMART; SM00345; HTH_GNTR; 1.
 SQ SEQUENCE 243 AA; 27190 MW; E90D8EAF2C2F2BD CRC64;

Alignment Scores:

Pred. No.: 10.4 Length: 243
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.88% Indels: 0
 DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9EV74 (1-243)

QY 344 CTACTTCTGCCAGGATAACCCTCGAA 370

Db 111 LeuThrSerAlaArgLeuThrLeuGlu 119

RESULT 12

Q9C196 ID Q9C196 PRELIMINARY; PRT; 411 AA.
 AC Q9C196;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CAMP dependent protein kinase regulatory subunit.
 GN PKAR.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ASPERGILLUS NIGER N400;
 RA Staudohar M., Bencina M., Van de Vondervoort P., Legisa M.,
 RA Panneman H., Rulijter G., Visser J.;
 RT "Cyclic AMP-dependent protein kinase is involved in morphogenesis of
 RT Aspergillus niger";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ296317; CAC36308.1; -.
 DR HSSP; P00515; 2APK.
 DR InterPro; IPR002373; CAMP_kin.
 DR InterPro; IPR000595; cNMP_binding.
 DR Pfam; PF00027; cNMP_binding; 2.
 DR PRINTS; PR00103; CAMPKINASE.
 DR SMART; SM00100; cNMP; 2.
 DR PROSITE; PS00888; cNMP_BINDING_1; 2.
 DR PROSITE; PS00889; cNMP_BINDING_2; 2.
 DR PROSITE; PS50042; cNMP_BINDING_3; 2.
 KW Kinase; CAMP.
 SQ SEQUENCE 411 AA; 44522 MW; BA43E6C393F09706 CRC64;

Alignment Scores:

Pred. No.: 9.7 Length: 411
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.70% Indels: 0
 DB: 3 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9C196 (1-411)

QY 2378 GTVGGTCTGTGGGCTGTGGGAGT 2352

Db 227 ValGlySerValGlyProGlyGlySer 235

RESULT 13

Q9K742 ID Q9K742 PRELIMINARY; PRT; 490 AA.
 AC Q9K742;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein BH3531.
 GN BH3531.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AF001519; BAB07250.1; -.
 DR InterPro; IPR003540; Binary_toxinA.
 DR Pfam; PF03496; Binary_toxinA.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 490 AA; 58355 MW; 2258B61391D47319 CRC64;

Alignment Scores:

Pred. No.: 9.47 Length: 490
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.70% Indels: 0
DB: 16 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9K742 (1-490)

QY 202 TCAGAAAGCTGAAAGATCTGGAAGGT 176

Db 398 SerGluargLeuylaspLeuGluGly 406

RESULT 14

O88831 ID O88831 PRELIMINARY; PRT; 587 AA.

DT 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE Ca+/calmodulin-dependent protein kinase beta (Cam-kinase kinase beta).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-CEREBELLUM;

RA Fujisawa H.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-CEREBELLUM;

RX MEDLINE=97420710; PubMed=9276695;

RA Kitani T., Okuno S., Fujisawa H.;

RT "Molecular cloning of Ca2+/calmodulin-dependent protein kinase kinase beta.";

RL J. Biochem. 122:243-250(1997).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AB018081; BAA33524.1; -.

DR HSSP; Q63450; 1A06.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00220; S_TKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 587 AA; 64446 MW; C9E49B72578F3971 CRC64;

Alignment Scores:

Pred. No.: 9.23 Length: 587
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 11 Gaps: 0

US-09-818-990B-1 (1-3963) x O88831 (1-587)

QY 2180 TCCCTCCAGCAGACCCAGCCGACCAA 2206

Db 571 SerProArgThrProGlnGln 579

RESULT 15

Q9LT87

ID Q9LT87 PRELIMINARY; PRT; 663 AA.

AC Q9LT87;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE Similarity to serine/threonine kinase.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=20277480; PubMed=10819329;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence

RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC

RL DNA Res. 7:131-135(2000).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AB025624; BAB02454.1; -.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR InterPro; IPR004040; STY_pkinase.

DR InterPro; IPR001245; Tyr_pkinase.

DR Pfam; PF00069; pkinase; 1.

DR PRINTS; PR00109; TYRKINASE.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00221; STYKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 663 AA; 73036 MW; 349C459F290A3C3B CRC64;

Alignment Scores:

Pred. No.: 9.08 Length: 663
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9LT87 (1-663)

QY 751 ACACGAGGTCTTCCCTTCATCTCTG 777

Db 211 ThrProGlySerSerProSerSerLeu 219

Search completed: November 30, 2002, 18:48:31

Job time : 144.5 secs

